For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex libris
universitatis
albertaensis

THE UNIVERSITY OF ALBERTA

**RELEASE FORM**

NAME OF AUTHOR Sherrie E. Shammass

TITLE OF THESIS An Experimental Investigation of Segment Duration and Intensity in English Juncture

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED Fall 1980

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

THE UNIVERSITY OF ALBERTA

# An Experimental Investigation of Segment Duration and Intensity in English Juncture

by

© Sherrie E. Shammass

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Science

IN

Speech Production and Perception

Department of Linguistics

EDMONTON, ALBERTA

Fall 1980

Digitized by the Internet Archive
in 2019 with funding from
University of Alberta Libraries

https://archive.org/details/Shammass1980

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled An Experimental Investigation of Segment Duration and Intensity in English Juncture submitted by Sherrie E. Shammass in partial fulfilment of the requirements for the degree of Master of Science in Speech Production and Perception.

## Abstract

Juncture has received attention in theoretical literature, but insufficient work has been done on this problem experimentally. In selected utterances which differ in the location of juncture, the duration and intensity of several segments were measured. Analysis showed that juncture location affects the duration and intensity of those elements close to the juncture. Perceptual experiments were designed in order to test whether these differences were psychologically important to listeners. Results indicated that these differences did play a perceptual role. Experiments also indicated that the response task did not affect listeners judgements to a marked degree. Other perceptual experiments were designed in order to test the effects of different combinations of segment durations. Results indicated that these effects are complex since silent portions of the signal may play a different perceptual role than speech sounds. Finally, models of juncture perception are tentatively proposed. Further experiments are also proposed in order to test several hypotheses.

## Acknowledgements

I would like to express my deep gratitude to my thesis supervisors, Dr. T. M. Nearey and Dr. J. Hogan for their constant guidance, abundant suggestions and generous assistance. Their continual aid and inspiration is greatly appreciated.

Many thanks are offered to the other members of my thesis committee, Dr. B. Derwing and Dr. K. Holden for their time and effort in reviewing this work and for their many helpful suggestions. I would also like to thank Dr. A. Rozsypal for giving me a better understanding of digital speech processing, an integral part of this thesis.

To my husband, Saeid, I express my deep appreciation for your constant support, encouragement, understanding, and shared enthusiasm for my work. I also express gratitude to my parents for instilling in me the worth of education and the love of knowledge.

# List of Tables

Table.....Page

1. ANOVA's for Duration ..... 43
2. Tukey Tests for Duration ..... 46
3. ANOVA's for Intensity ..... 49
4. Tukey Tests for Intensity ..... 50
5. Classification Results of Discriminant Function Analysis ..... 52
6. Subjects' Responses for the Forced Choice Experiment 62
7. ANOVA Results for the Crossed Experiment ..... 97
8. Logistic Results for the Crossed Experiment ..... 106
9. Backwards Stepwise Analysis ..... 109

## List of Figures

Figure.....page

1. Block Diagram of Stimulus Preparation ..... 39
2. Subject-Juncture Interaction ..... 45
3. Scatterplot of Disriminant Function Analysis ..... 53
4. Block Diagram of Stimulus Presentation ..... 61
5. Identification Curves for the Forced Choice Experiment ..... 63
6. Identification Curves for the Free Choice Experiment 66
7. Indentification Curves for the Free Choice Experiment (for subjects responding with all three categories) ..... 68
8. Swamping Curves ..... 73
9. Crossed Experiment Curves ..... 84
10. S#S Interaction Plots ..... 99
11. K#S Interaction Plots ..... 101
12. S#N Interaction Plots ..... 103

Table of Contents


Chapter Page

I. Introduction .....1

II. Historical Review .....6

A. Formal Investigations .....6

Summary .....16

B. Experimental Investigations .....18

Acoustic Phonetic Studies .....18

Physiological Studies .....25

Perceptual Studies .....27

III. Measurement Study of Production .....35

A. Description and Measurement of Data .....35

Speakers .....35

Materials .....35

Apparatus .....36

Recording .....37

Digital Gating .....38

Segmentation and Measurement .....38

B. Statistical Analysis: Results and Discussion ...41

Analysis of Variance for Duration .....42

Tukey Tests for Duration .....44

Analysis of Variance for Intensity .....47

Tukey Tests for Intensity .....48

The Discriminant Function Analysis .....51

Summary .....54

IV. Identification Tasks .....55

A. Experiment 1: Identification by Forced Choice

Method .....56
Preparation of Stimuli .....56
Lists of Stimuli .....58
Subjects .....59
Apparatus .....59
Procedure .....60
Results and Discussion .....60
B. Experiment 2: Identification by Free Choice Method .....64
Subjects .....65
Stimuli and Apparatus .....65
Procedure .....65
Results and Discussion .....65
V. Combinatory Perceptual Studies of Juncture .....69
A. Experiment Three: Swamping Experiment .....70
Subjects .....70
Lists of Materials .....70
Apparatus .....71
Procedure .....72
Results and Discussion .....72
Rating Results .....79
B. Experiment Four: The Fully Crossed Experiment ..81
Subjects .....81
Lists of Stimuli .....82
Procedure .....82
Results .....83
Statistical Analysis .....95
VI. Perceptual Models and Discussion .....112

A. Some Bottom-up Models of Speech Perception ....113

Models from Research in Computer Speech Recognition .....117

B. Models for Juncture Perception .....118

No Conflict in Cues .....120

Conflict in Cues .....127

C. Discussion .....131

Conclusions .....134

Bibliography .....136

APPENDIX 1: Stimulus Sentences .....143

APPENDIX 2: Means and Standard Deviations of Measurements .....144

APPENDIX 3: Duration and S Intensity Values .....145

APPENDIX 4: Segments Used in the Crossed Experiment .....146

# I. Introduction

Linguistic investigators had long ago noted that word boundaries, or junctures, play an important role in the sound patterns of language. Minimal pairs such as "an aim" vs. "a name" were considered to differ only in one aspect - the placement of juncture. Much controversy emerged, however, over the role and nature of juncture. Earlier inverstigators argued about how juncture should be treated in the phonology of a language. Should it be considered a separate phoneme? Does it have allophones? Or is it a suprasegmental feature? Such questions prompted much theoretical speculation on the nature of juncture that was often dependent entirely on the general linguistic framework at hand.

Since juncture played so important a role in descriptive systems, an important question arose regarding its physical correlates. Is there a phonetic basis for juncture? In other words, is juncture based on real physical events from the speech signal, or is it merely a psychological entity imposed by linguistic perception? Investigators thus attempted to define physical correlates of juncture in terms of both production processes and acoustic events. Recently, questions regarding the perception of juncture have emerged. Is the perception of juncture related to the physical manifestations of juncture? More importantly, what particular physical elements in the speech signal cue the existence of juncture to the

perceiver? Such questions have been addressed by recent experiments, but the area of juncture perception is still relatively unexplored.

This work will concentrate on the perception of juncture based on more or less continuous aspects of segmental elements of the speech signal. In particular, duration and intensity of clusters involving the fricative /s/ with the sonorants /n/, /m/, /l/, and /w/ are explored in relation to both the production and perception of junctural placement. It was decided that an investigation of the fricative's junctural behavior was appropriate for several reasons. First, /s/ is not associated with different spectral patterns depending upon junctural presence, nor does it contain qualitatively distinct 'marginal' or boundary allophones, e.g., such as aspiration or devoicing, as in the case of stops. Thus, an investigation into the effects of the continuous properties of duration and intensity is feasible. Second, /s/ is a prime element in English clusters, and therefore, clustering effects could be studied. Third, /s/ is an important morphological suffix, associated with meaningful semantic changes (e.g. third person singular and plural markings). Thus, subject identification of juncture could be based on meaningful differences. The sonorants were examined because their behavior within clusters at boundary points are largely unknown; as well, most of these elements(/n/, /m/, and /w/) are not associated with major allophonic differences near

the junctures to be examined.

The following chapter outlines past investigations of juncture phenomena. The first section reviews theoretical and formal arguments within several general frameworks. It is noted that arguments are largely based on the investigator's own intuition, and the choice criteria involve formal elegance and logic. The second section reviews experimental research in juncture phenomena. Measurement studies in both production and acoustics are reported. Finally, previous perceptual experiments are reviewed. It is noted that there have only been few experimental studies in the area of juncture phenomena and that this area requires more empirical study.

Chapter Three outlines a measurement study of production, where differing juncture placements between /s/+sonorant strings are produced. A second factor is contrastive stress, which, in one condition is placed on the pre-junctural word, and in the other condition placed on the post-junctural word. The effects of contrastive stress, clustering, and change of word boundary are measured, and their interrelationships are discussed. Both intensity and duration differences are recorded. These measurements are then used in delimiting the changes done to signals for the purposes of speech perception studies.

Chapters Four and Five outline four perceptual studies. Preliminary identification experiments are outlined in Chapter Four; more elaborate experiments are in Chapter

Five. The first study consists of forced-choice identification of juncture placement in typically produced frames. This study produced a 'base-line' of identification curves and furthermore, showed that duration differences did, in fact, alter listeners' choices. The second study was a 'free-response' identification of these frames. Results of this investigation indicated that the response task did not affect listeners' choices to a marked degree. However, there were subject differences involved, indicating some differences in listener strategies. The third study tested the effects of altering one element in a typical frame. It was found that the /s/ contributed to large perceptual changes, but the other elements (/k/, pause, and /n/ duration) merely strengthened or weakened the frame response. The fourth study was the major investigation involving crossing possible duration levels, in a forced-choice task. Results indicated that all elements have an affect on listener's perception, but it was not fully clear, given the nature of the data, how these elements interacted.

Chapter Six discusses the possibility of a 'bottom-up' model of speech perception. This chapter outlines several bottom-up models, and discusses their merits and difficulties. Several models based on categorization data are reviewed: the 'context allophonic', the 'diphone', the 'whole-word', the 'feature detector', the 'memorial', and the 'allophonic detector' models. Two segmentation models

are proposed: the 'segement-allophone' and the 'pairwise-evaluation' models. It is hoped that models of this kind and previously proposed models of 'top-down' processing can eventually merge for a comprehensive model of juncture perception. Implications for future research are mentioned in this regard.

## II. Historical Review

### A. Formal Investigations

"It was observed quite early that there is no one to one correspondence between grammatical words - normally written between spaces - and the internal structure of a spoken chain of sounds" (Lehiste, 1960. p.5). How then, were word divisions to be treated in phonological descriptions? Sweet (1913) suggested that the spaces between words in orthography should be abolished and instead, the 'phonetic word' be considered as the basic unit of speech. That is, the internal structure of a spoken chain of sounds should dictate word divisions. Lists of minimal pairs differeing only in the placement of juncture were compiled by Jones (1931). Trubetskoy (1935, translated 1969) was perhaps the first to theorize on the nature of these differences; he termed the phenomena 'Grenzsignale' (boundary signals), but considered them largely 'non-phonemic'. Rather, he likened them to traffic signals which keep the flow of speech understandable by segmenting sentences, words and morphemes. Bloomfield (1933) suggested that juncture phenomena were simply a subset of stress phenomena, and said that minimal pairs differing in juncture placement really differed in the moment at which loudness begins to increase.

The first comprehensive placement of juncture phenomena into a phonemic system was accomplished by Trager and Bloch (1941), and later elaborated by Trager and Smith (1951).

These authors defined two major types of juncture - open and closed. Open juncture is the transition between a pause (either at the end or beginning of the utterance) and the segmental phoneme bordering that pause. More explicitly, it is the sum total of features of the phonemes bordering the pause. A distinction was made between 'internal open juncture' and 'external open juncture'. Internal open juncture refers to features of open juncture that are present within an utterance, and is symbolized by hyphens. External open juncture refers to features of open juncture that are present across utterances, and are symbolized by spaces between segments. Closed juncture is the transition between two segmental phonemes in the same utterance, and is "taken to be the manner of normal transition from one phoneme to another". The authors symbolized close juncture implicitly by writing the segmental phonemes without a space, and explicitly by a 'tie-line' (‿) between the two segments. However, they stress that the 'tie-line' is non-phonemic. In the 1951 refinement of the system, junctural, stress and intonation patterns were given phonemic status. Internal open juncture was represented as the phoneme /+/; thus, for example, the difference between 'nitrate' and 'night rate' was considered as a difference of transition between /t/ and /r/, and was thus represented as /naytreyt/ and /nayt+reyt/, respectively.

The controversy over the nature of junctural phenomena was initiated by Moulton's analysis of German (1947). He

postulated allophones of juncture. According to Moulton, open juncture is a segmental phoneme having two allophones: at the bounds of an utterance it is manifested by a pause, while within the utterance it appears either as a brief pause, or is in free variation with 'zero' (/ϕ/). Leopold (1948), however, pointed out that the zero allophone only occurs when there is word-formation processes (e.g. compounding, suffixation) and that, therefore, there is morphophonological conditioning involved. Pike (1947) argued that in fact, the morphology involved should be taken directly into consideration when postulating the existence of juncture. Thus, he argued that junctures are bound not only to the surface phonetic forms, but also to the morphology of a language.

Harris (1951), on the other hand, introduced juncture "as a factor in phonemicization, but only, of course, to the extent that this is possible without knowledge of morphemes". Juncture, for Harris, is a phoneme set up for the purposes of simplifying the structural description by reducing the number of phonemes. He outlined a specific procedure for the introduction of junctures in a phonological description. This basically entailed altering the environment of one phoneme set by postulating the existence of a juncture, so that its environment is non-identical with its counterpart phoneme set. For example, the /ay/ of 'minus' and the /Ay/ of 'slyness' seemingly need two phonemes, /ay/ and /Ay/ because a meaningfull contrast

is obtained in the same environment. However, Harris suggested that the /Ay/ of 'slyness' be represented as /ay-/, where the hyphen represents a juncture phoneme. Since the two phonemes /ay/ and /Ay/ can now be represented as /ay/ and /ay-/, only one phoneme, /ay/ is needed. Thus, two sets of phonemes are replaced by one set. Similarly, the unreleased set of stops, /p'/, /t'/, /k'/, can be represented as /p/+/#/, /t/+/#/, /k/+/#/ respectively, thereby reducing the two sets of phonemes (/p/, /t/, /k/ and /p'/, /t'/, /k'/) into the one set /p/, /t/, /k/. Harris' juncture is a 'zero phoneme', having no phonetic properties of its own, but rather, acting strictly as environment for the purposes of reducing the phonemic inventory.

In the same vein, Chomsky, Halle and Lukoff (1956) argued that junctures are not in themselves, physical entities, but are rather introduced in order to reduce the number of features that must be considered phonemic. However, these authors do equate juncture placement to morpheme boundary location; junctures are only placed in the system where phonetic effects are correlated with morphemic boundaries. Furthermore, different junctures can represent different morphological and syntactic constructions. Thus, Pike's original argument (1947) of the infiltration of higher level systems (ie. morphology and syntax) into the phonemic system was recalled.

Hockett (1955), in contrast to Harris' non-physical phoneme of juncture, set up a phoneme of juncture which is

associated with all types of physical changes which occur at boundaries. Thus, allophonic changes at boundary points are discussed in terms of the presence of the juncture phoneme; different classes of allophones are merely 'allo-junctures'. Of course, allo-junctures are not phonetically homogeneous, but Hockett considered this a benefit rather than a setback.

> "It takes longer to describe all the allophones of a juncture, but once the juncture has been described, it constitutes a powerful tool... [T]he heterogeneity of the allophones of a juncture renders possible a much neater phonetic layout for 'ordinary' phonemes, because one has drawn from the latter many of the messy marginal differences which would otherwise yield a very complex system." (Hockett, 1955, p. 171)

Hockett points out that, although the phonetic manifestations differ, the structural consequences are homogeneous. Furthermore, Hockett clearly states that juncture phonemes are not related to higher grammatical boundaries, such as boundaries between words. For example, the homophony of 'finder' and 'find her', according to Hockett, displays different word boundaries, but not different junctural boundaries (Hockett, 1958).

Hill (1958) provided a new approach to the treatment of juncture, by considering it a suprasegmental, timing phenomenon. According to Hill, the distinctions between, for example, *that stuff* and *that's tough* is chiefly in the prolongation of the final /t/ of 'that' in the former vs. prolongation of the /s/ in the latter. Furthermore, he considered that these prolongation differences were only 'half-units' i.e., half the time of a 'normal' segmental

sound. Hill discussed the possibility that the allophonic changes surrounding a boundary were redundant in light of the redefinition of juncture phenomena in terms of time phenomena. In a footnote, Hill also raised the possibility that intensity differences may be involved, a point which he states "will have to await acoustic analysis". It was only two years later, in 1960, that the first published experimental work on the subject emerged; Lehiste experimentally measured the acoustic manifestations of juncture. Her work is more fully described below.

The trend towards experimental analysis of junctural phenomena was set aside due to the increased attention paid to the new Transformational Generative Grammar (TGG) framework. Chomsky and Halle (SPE, 1968, p. 369) stated that acoustic phonetic factors need not be taken into consideration: "the requirement of phonetic effects of some sort be associated with word boundary appears as insufficiently motivated, and we have not incorporated it into our theory of language." They give no substantial reason for this decision, other than a shaky comment regarding the 'competence-performance' distinction. In SPE, boundaries thus become more abstract and more tightly connected to the higher level grammatical structure of the language. In fact, the top-down machinery of TGG supplies the boundaries (the boundary # is inserted automatically at the borders of each string dominated by a major category). These boundaries are then modified by rules to obtain

correct output. Economy and generality are the main considerations. For example, the stress rules of English are greatly simplified, but this is done at the expense of positing abstract boundaries in the underlying representation. Boundaries, in the Chomsky and Halle sense, are considered as "units in a string... on a par with segments". Thus, like segments, boundaries are "complexes of features". Phonological rules are simplified, in that they may apply to strings which contain specific types of boundaries. Thus, boundaries are also used for the purpose of blocking specific rule applications, which in turn, allow rules to overgeneralize. At the end of the grammar, junctures are erased.

McCawley (1968) retained much of the Chomsky and Halle approach, and further refined the system. Following the SPE model, he proposed that a hierarchy of junctures be set up, implying that there is a fixed order relationship between different types of junctures. He also suggested that a 'rank' of a rule be used in order to indicate the scope of a particular juncture. Thus, for example, a string such as #A:B#C#D:E#, where # and : are junctural elements, and # is of a higher rank, would be segmented into AB,C,DE by a #-ranking rule, and into A,B,C,D,E, by a rule of rank :. Therefore, the hierarchic classification makes all #'s also count as :'s. In McCawley's system, junctures are assumed to occur at morpheme boundaries only. Therefore, he proposed that the morpheme boundary (symbolized by &) be the lowest

juncture in the heirarchy. The highest juncture, he said, is a pause, symbolized as $. Furthermore, a rule becomes more complex by a junctural element only when that element is lower than $ (i.e. it is 'marked'). Finally, McCawley allowed juncture insertion rules in addition to supplying junctures by general syntactic conventions. Therefore, a rule of the type $\emptyset \longrightarrow$ #/N_ was permissable. He completed his outline of junctural elements by stating a few ad hoc conditions on the interaction of junctures and transformations, such as "a constituent transportation also carries the junctures" and "a deletion rule leaves behind the stronger of the junctures which had been at the borders of the deleted constituent" (1968, p. 58).

McCawley's ideas were extended by Harms (1968) and Stanley (1973). Harms outlined a feature matrix system which portrayed McCawley's hierarchy in more depth, while Stanley attempted to classify types of interactions between boundaries and phonological rules in terms of the hierarchy of boundary types. Like McCawley, Stanley argued that the 'stronger' boundary type always 'won' at each stage of the grammar. However, Stanley also introduced a notion of 'boundary weakening', whereby initially assigned #'s at the borders of stems were weakened, depending upon the class of affix to which the stem was attached. These classes were hierarchically organized on the basis of the affix's phonological abilities to combine with adjacent materials.

Stanley's paper paved the way for other types of

boundary mutation. Not only boundary weakening, but also boundary strengthening was invited (cf. Selkirk, 1974 and Sag, 1974). Eventually, all types of boundary changes found their way into the grammar (eg. + $\longrightarrow$ #, ## $\longrightarrow$ #, etc.). These boundary mutations were designed to handle the problems of 'exceptions' (particularly with morphological exceptions). As a consequence, exceptions, which were previously taken care of in the lexicon (i.e., specially marked), instead became handled differently in the course of derivation by rules of boundary mutation. Again, the criterion of generality evoked this treatment. In other words, generalizable exceptions (or classes of exceptions) were handled by rules, where these rules involved the changing of boundary assignment.

The concept of boundary mutations grew more extensively with the introduction of other generative concepts, such as context sensitivity (Sag, 1974), rule ordering (cf. Devine and Stevens, 1976), and universal strength hierarchies (Lass, 1970). Thus, the 'machinery' for boundary mutations grew to the point where abstract generalizations became the norm. As with generative systems in general, the system for treating boundaries became too powerful, too abstract, and quite misleading.

It became the burden of 'Natural Generativists' to constrain the abstractness and use of boundaries. They proposed that boundaries could not be presented in the underlying form unless motivated by some morphological

behavior. This contrasts with the generative approach, whereby abstract boundaries were motivated for the purposes of generality and simplicity; it also contrasts with the structuralist approach, whereby boundaries were motivated by phonetic conditioning. The natural generativists distinguished between the various types of phonological rules; thus, they made a distinction between 'true' phonological boundaries and boundaries which were specified in terms of the morphology. As a consequence, the number of different types of boundaries decreased, while the 'exception' lexicon increased. Indeed, the lexicon, in this approach, now had to handle exceptions, as well as problems left over from the abolition of rule ordering. The 'machinery' of the grammar was reduced, but the lexicon swelled.

Hooper (1975) exemplified the NGG (Natural Generative Grammar) approach; she stated that word boundaries were permissable in morphological rules, but not in purely phonological rules. Anderson (1974) made a distinction between morphological, phonological, and phonetic rules. He stated that phonological rules could include boundary elements (and also reference lexical class), while phonetic rules could only reference "phonetically realizeable boundaries". Rhodes (1974) eliminated boundaries from his natural processes by "transderivational constraints", which were basically constraints on derivations containing the same morphemes. He then had to specify the location of

syntactic and morphemic boundaries by indirect means. Basically, a segment was considered to be adjacent to a boundary if there existed a derivation of the morpheme containing that segment, such that the segment occured next to silence. This was nothing more than bringing the old structuralist idea that a boundary existed if there was a 'possible' or 'optional' pause. Vennemann (1974) replaced boundaries with a convention of marking 'consonant strength' in the lexicon; Vennemann's lexicon contained "words rather than stems" and basic allophones were directly listed in the lexical representation. Valid generalizations of allophonic variation were expressed by means of redundancy rules.

This controversy over the presence of boundaries in phonology led Devine and Stephens (1976) to review their usefulness in the system. These authors concluded that

> "...a grammar can produce correct outputs without phonological boundaries simply by a direct correlation of the rules of the phonology with the boundary sequences of the morphosyntax. Phonological boundaries capture generalities about correlations between phonological rules and morphosyntactic boundary sequences. Other organizations capture certain similar generalizations, but phonological boundaries capture some generalizations that all other organizations fail to capture, and there are no generalizations captured by other organizations that cannot also be effectively expressed in terms of phonological boundaries."

**Summary**

The role of boundary phenomena in phonological descriptions has varied considerably, depending largely on the basic descriptive framework involved and related definitional differences. Pre-structuralists maintained that

boundary phenonomena existed, but that they were not to be treated separately from phonemic theory. Structuralists included boundary phenomena in the phonological system as phonemes; however, they maintained that these phonemes must be identifiable from the surface data. Generativists kept the boundary phonemes in their system, but maintained that they need not be manifested in the surface data, but could be more abstract and motivated by morphological and syntactic considerations. Natural generativists argued that boundaries in phonology must either be surface-conditioned, non-abstract entities or else abolished altogether in favor of more complex lexical representations.

These shifts of definition of the term 'boundary' in phonology have led to various forms of phonological systems. Thus, the form of the lexicon and rules differ in these systems, depending on the use and definition of the boundary component. The various phonological systems that have been discussed thus far differ as to the form and placement of boundaries. A more compelling and fundamental question, however, is whether boundaries even exist at all, as opposed to how they exist or where they exist. Cena (1978, p.1) puts the issue this way:

> "Competing rules (or definitions) within theories may differ crucially only in their form...However, the question of the substantive reality of a rule (phenomena) takes precedence over the question of its formal validity since it is pointless to examine the correctness of a form of a rule (phenomena) if the (phenomena) in the first place has no substantive basis".

This more important question was not addressed by the

aforementioned theoreticians who all implicitly assumed the existence of 'boundaries'. However, it was addressed by experimentalists who intended to explore the acoustic, physiological, and psychological manifestations of so-called 'boundary phenomena', or juncture.

## B. Experimental Investigations

There have been several experiments investigating the nature of juncture. Since juncture phenomena entails the study of allophonic variation, studies which have investigated juncture are based on data describing allophonic distinctions. There have been three main types of experimental studies on juncture: those dealing with acoustic output, those dealing with physiological production, and finally, those dealing with perception of juncture.

### Acoustic Phonetic Studies

The earliest studies which experimentally investigated juncture phenomena were acoustic phonetic studies. These studies typically investigated the acoustic manifestations of boundary presence.

The first study accomplished was that done by Lehiste (1960). She was interested in measuring and defining the changes which take place phonetically when a word boundary is imposed. Her study used minimal pairs differing only in juncture placement, such as "a nice man" vs. "an ice man". She selected those minimal pairs which were highly

identifiable as distinct utterances, and then identified junctural cues on the basis of sonographic measurements of differing aspects between the two utterances. A summary of her results follows:

1. there exists glottalization (glottal stop) or laryngealization for word initial vowels;
2. final vowels are very long in duration;
3. initial nasals are longer than final or medial nasals, and final nasals are longer than medial nasals;
4. /s/ durations are longer in both final phrase position and initial word position;
5. stop durations are longer in initial position;
6. voiceless stops have an aspiration cue in initial position;
7. final /l/ is longer than initial and medial /l/; formant differences appear betwen /l/'s in different positions of the word;
8. intensity differences are often apparent; normally, initial allophones increase in intensity while intensity decreases in final allophones. However, this pattern is inoperative in the case of voiced stops.

Lehiste concluded by refering to 'bound' utterances in which there exist initial and final allophones which cue junctures. She also suggested that the most prominent cues which aid listeners seem to be largely segmental in nature (i.e., aspiration, glottal stop placement, etc.), whereas suprasegmental cues such as intensity aid the listener, but

less so than segmental cues. However, duration is also considered a primary cue, since differences of duration affected juncture perception.

Lehiste (1964) also conducted studies investigating the allophones of /r/ and /l/. She found that some allophones of /r/ have reliably stable formant patterns which signify boundaries of larger phonological units. For example, initial /r/'s have relatively low formants (F1 is less than 300 Hz, F2 is less than 1000 Hz, F3 is less than 1400 Hz) while /r/'s in all other positions have much higher formant values (e.g., syllabic final /r/ has F1 less than 450 Hz, F2 less than 1400 Hz, and F3 less than 1600 Hz). Allophones of /l/ are also marked by distinct formant patterns. The final allophone of /l/ has a clearly defined acoustic structure which is not influenced by the preceeding vowel; the second formant anticipates the second formant of the following vowel. For intervocalic /l/ allophones, F1 and F3 vary according to the first and third formants of the preceeding vowel, and F2 varies according to the second formant of the following vowel. Morphologically significant /l/-type endings were distinguishable from one another; for example, for the '-ly' in "solely" and "cooly", differences in the vowel /i/ were found.

Hoard (1966) verified that the phonetic correlates of juncture, as defined in the Lehiste study, were maintained in connected discourse. He used four native English speakers, and eighteen listeners, and analyzed correctly

identified choices from juncture minimal pairs by spectrographic means. Hoards' data indicated that segment duration is a "systematic acoustic correlate of juncture" while fundamental frequency and amplitude are not. In addition, allophonic distinctions indicating junctural presence were maintained in connected speech.

Lisker (1965) conducted a measurement study of /s/-stop sequences in which the boundary was varied (eg. /s/+/t/, /s//t/+, /s/+/s//t/, +/s//t/). He measured both the /s/ frication and the stop closure duration. His data further verified Lehiste's results; for both /s/ and the stops, final phonemes were significantly shorter in duration than initial phonemes.

A more recent study measuring allophonic variation was conducted by Umeda and Coker (1975). This important investigation revealed that sub-phonemic details were consistently manifested in production; Umeda and Coker classified these variations in terms of simple rules. Their study covered variations of consonant durations, and the allophones of voiced and voiceless stops, as well as variations of the other consonants found in American English.

These investigators found that consonant duration varies according to stress, position and context. The conditions which influenced consonant durations are not interactive, however, and can be explained by a linear additive model. Two major rules were proposed: the

lengthening factors are stress, word boundaries, and pauses, whereas the shortening factor consisted of being adjacent to a fricative. The lengthening factor of "presence of boundary" was further detailed: "Boundaries between function words do not lengthen consonants; boundaries with strong content words lengthen consonants". Thus, word importance is also a factor in consonantal lenghtening.

The duration of stops and nasals is also influenced to a high degree by contextual factors. Preceeding and following consonants affect their durations, even across word boundaries, but stress does not affect stop and nasal durations as much as it affects fricative durations. The authors suggested that segmental allophonic variation plays a greater role in stops, but durational allophonic cues are important in fricatives (which have little segmental allophonic variation).

Umeda and Coker (1975) found that allophonic variation was characterized by devoicing time for voiceless stops, vocal cord oscillation for voiced stops, and amplitude differences for other consonants. Initial word and initial stressed voiceless stops are marked by aspiration and have a burst. Voiceless stops which are positioned near nasals, or which are in function words, are shorter and less aspirated. Final voiceless stops in vocalic contexts are unaspirated and have no burst. (An exception is found with /t/, which becomes aspirated when a non-sibilant consonant preceeds it. Prepausal /t/'s have approximately 30 msec. of aspiration.)

For voiced stops, two aspects are allophonically important - intensity and harmonic quality. These factors can be discussed in terms of vocal cord oscillations. Basically, word and stress-initial voiced consonants are lower in intensity than medial and final voiced consonants. However, when Coker and Umeda (1975) attempted to synthesize this distinction, they found that amplitude control was done more effectively by lowering the first formant of initial allophones. Thus, they suggested that the contrast between initial and non-initial voiced consonants lay in spectral differences in the voice bar. These differences could be caused by either the excitation source or by vocal tract transmission (i.e., either the source or the filter was causing these differences). From glottographic data, word-initial voiced consonants were found to have higher glottal amplitude oscillations. The authors suggested that the excitation source was causing the difference in spectral bar detail for the following reason: from spectrogram analysis of the voice bar, initial stressed voiced stops have strong fundamental frequencies and lack higher components, but stressed consonants are produced with tenser cheek muscles than unstressed consonants which would have the effect of raising the fundamental, rather than lowering it. Therefore, the difference between initials and non-initials must be due to the glottal source.

Using a fiber-optic technique, Coker and Umeda (1975) found several differences of vocal cord oscillation between

initial and final voiced consonants. These differences result in differences of spectral quality. Basically, for initial consonants, the vocal cords gradually come apart at the transition from the preceeding phone to the voiced consonant. The cords continue to spread until the closure period of the stop, when they close slightly, or even incompletely close; complete closure is not attained until a few pitch periods after the start of the following vowel. Thus, word initial voiced consonants have the following properties:

1. the intensity of glottal vibration is high;
2. there is no closure of the glottis;
3. the glottal spectrum is mostly fundamental;
4. the intensity of the fundamental is high;
5. the intensity of the harmonics is low; and
6. the glottal waveform is sinusoidal.

For word final voiced stops, on the other hand, the glottis closes completely, with the result that the glottal waveform produced is not sinusoidal, but rather sawtoothed, and therefore is richer in higher harmonics. Therefore, word-final voiced stops have the opposite properties than those found for word-initial voiced stops. Also, initial voiced consonants have plosive bursts, while final ones do not (i.e., final voiced consonants are 'devoiced').

For other consonants, peak-to-peak amplitude patterns parallel the devoicing time of voiceless stops. Consonants in word initial positions have amplitude distributions which

are fairly sharp, while word-medial or final consonants have spread amplitude distributions. For /n/, the intensity is not constant, but tends to increase in time in initial position. Voiceless fricatives have high amplitudes in initial and stressed positions, but low amplitudes in medial and final positions; prepausal voiceless fricatives have the lowest amplitude. However, voiced consonants, which clearly have segmental cues for differentiating word-initial and word-final allophones, have lower amplitudes in initial position than in final or medial positions.

Coker and Umeda's study dealt with the acoustic consequences of physiological happenings which occur at word boundaries. The following section will describe studies which investigate some important aspects of the physiological articulation at junctures.

**Physiological Studies**

Studies of the articulation near junctures typically involve the notion of "coarticulation blocking". Context effects on articulations are well known; anticipation of the following phone and the effects of a previously articulated phone affect the articulatory behavior of the phone which is being articulated. It was hypothesized that coarticulation would be blocked at juncture boundaries.

McLean (1973) studied forward coarticulation effects of velar movements at junctural boundaries. Using cineradiographic techniques, he found that velar coarticulation was delayed consistently where a marked

juncture boundary existed. McLean kept the lexical items constant and varied the type of juncture (as defined by Trager and Smith) within these words, which were placed in different sentence frames. A "marked" juncture boundary typically contained prosodic markings to indicate the existence of the boundary. Thus, McLean actually found that coarticulation is blocked at sentence or clause boundaries, and not, per se, at lexical or word boundaries.

Lewis, Daniloff and Hammeberg (1975) investigated the articulatory aspects of coarticulation in word-level situations. They studied the dental allophone of /n/ in /nθ/ and looked at dental coarticulation which was indicated by dental contact during the pronunciation of the /n/. They found that coarticulation effects were not impeded by most junctures; in fact, coarticulation was blocked only in the presence of long pauses. The authors felt that coarticulation is therefore not useful in marking juncture. Their results also seemingly conflict with those of McLean (1973); however, this is not the case, since McLean's results were based on prosodically marked junctural boundaries, and not word-level juncture.

Bladon and Al-Bamerni (1976) looked at allophonic variation in terms of coarticulatory resistance. They studied spectrograms of the different English /l/'s and found clear /l/ (word initial /l/) admits an exeptionally large degree of coarticulation, dark /l/ (word final and word medial /l/) admits a less amount of coarticulation and

syllabic /l/ (word final /l/) is very resistant to coarticulation effects. These results agree with Lehiste's original study (1964) on /l/ allophones. However, the notion of coarticulation blocking is more properly defined within the scope of articulatory behavior, rather than in terms of acoustic output.

**Perceptual Studies**

How is the acoustic output of juncture percieved? Of the cues mentioned above, which are important for listeners?

Christie (1974) used synthetic speech in studying the cues for juncture perception. The author was interested in syllabic juncture, and used the nonsense word /asta/, which could be varied in order to perceive either /as-ta/ or /ast-a/. Christie found that formant transitions have little or no effect on the boundary, but silence and aspiration both affect the location of where the boundary is perceived.

Similarly, Darwin and Brady (1975) examined the cues which distinguish the phrases "I made rye" vs. "I may dry", using real speech. They found that if the stop closure is short, formant transitions have little effect on the boundary.

Some other important studies have used real speech in order to find the perceptual cues involved in juncture perception. (The use of real speech offers direct implications for perceptual cues, whereas synthetic speech is often plagued by the uncertainty as to whether or not its synthetic nature is altering perceptual responses.) Several

other juncture studies using real speech have been done: some have dealt with the locus of segmental cues for word juncture, others have investigated suprasegmental aspects of juncture perception, and still others have studied some durational cues associated with juncture.

Nakatani and Dukes (1977) perfomed a listening experiment using minimal pairs differing in junctural placement (i.e., similar to Lehiste's minimal pairs). By splicing parent minimal pairs in varying locations, they obtained a full set of offspring utterances which contained varying amounts of particular juncture cues. Listeners judged which of the two parent phrases the "new" utterance more closely resembled. They found that the cues for the perception of word juncture occured at the word offset only for /l/ and /r/, and at the word onset for all other phones, but never occurred medially. That is, listeners perceived hybrids containing more junctural cues at word onset as though it were from the parent minimal pair which contains those cues; more medial cues did not affect listener's choice of parental utterance; and more final cues only affected listener's choices for those words ending in /l/ and /r/. The authors suggested two simple perceptual strategy rules:

1. what we hear at the beginning of the word depends on how the word begins, but
2. what we hear at the end of the word depends on how the next word begins, except for cases involving /l/ and

/r/.

Spectrographic studies done by Nakatani and Dukes showed that the cues most important for change of juncture perception were burst, aspiration, glottal stop placement, laryngealization, and distinct syllable initial allophones of /l/ and /r/. Adding more durational information into the hybrid utterances did not affect listeners' judgements regarding juncture placement, and hence, the authors concluded that duration is not a juncture cue. However, they admit that their choices of minimal pair utterances could have been a factor, since all pairs could be differentiated solely on the basis of segmental allophonic cues; no pair was tested which differed only in allophonic durational differences.

The prosodic cues for word perception were investigated by Nakatani and Schafer (1978). They gave listeners utterances in which two adjacent words from the utterance were replaced by corresponding numbers of /ma/ syllables. For example, if the original utterance was "The remote stream was perfect for fishing", listeners heard "The mama ma was perfect for fishing". The stress pattern, as defined by stress transcriptions of the signal by trained phonetic listeners, was altered according to the intended phrase substituted by /ma/'s. Listeners were instructed to parse the 'mama' phrases according to whether the first word had one or two syllales (i.e., ma/mama or mama/ma). The authors found that most (but not all) subjects could correctly parse

the utterances when these differing stress patterns were given. That is, most subjects could tell whether a monosyllable was followed by a bisyllable or the reverse, on the basis of stress pattern alteration alone. For those phrases with ambiguous stress patterns, subjects were still able to parse the utterances at a better than chance level. Therefore, the authors concluded that other prosodic features were also involved.

Using the technique of hybrid speech synthesis (Olive and Nakatani, 1974), which is a technique designed to assess the strength and interactions among speech features, Nakatani and Schafer studied the effects of rhythm, pitch and amplitude. Rhythm patterns were hybridized by linearly stretching or compressing the /m/ and /a/ segments, in synchrony with other alterations of pitch and amplitude so that naturalness was retained. The authors found that alterations of rhythm affected listeners' parsing responses, but that alterations in pitch and amplitude did not affect listeners' performance. They then tested what specific attribute in rhythm was important by measuring durations of the syllables in the /mama/ utterance. The authors found that words judged as monosyllabic were longer than comparable syllables in a word judged as being bisyllabic, and that word-initial consonants were also longer.

Nakatani (1979) suggested that both allophonic and prosodic cues exist for parsing speech. He suggested that there were two basic kinds of parsing cues, which he called

fission and fusion cues. Fission cues indicate that the utterance must contain a boundary. These include allophonic variation cues (e.g., aspiration of word-initial voiceless stops), and prosodic stress and rhythm cues (e.g., consecutive primary stressed syllables must belong to different words). Fusion cues indicate that the utterance cannot contain a boundary and include allophonic variation (such as syllabic nasal), and prosodic stress and rhythm cues (such as an unstressed syllable in a content word). In Nakatani's parsing experiments, using utterances which contained both fission and fusion cues, listeners parsed on the basis of both these cues.

Nakatani and Dukes (1979) compared allophonic, stress, and prosodic cues for word perception. They used phrases containing words with more than one syllable, such as "maid enforced" vs. "maiden forced". They found that allophonic variations were the most salient parsing cues, stress cues were salient to a lesser degree, but prosodic cues such as pitch and rhythm were not salient.

It is important to note that in Nakatani and Schaffer's 1977 experiment where the segmental, spectral allophonic cues were eliminated, rhythm was found to be a parsing cue, but when these stronger allophonic cues were included in his 1979 study, rhythm was no longer found to be a parsing agent. It is also important to note that in the 1977 experiment, the syntax remained constant, which is not the case in the 1979 study. Many of the minimal pairs in this

later study differed in the location of the major syntactic boundary.

Andresen (1979) made a similar mistake in his identification experiments. Listeners were asked to correctly identify intended utterances, which had counterpart utterances in juncture location (e.g., s*yntax* vs. t*in-tax* vs. fl*int-axe*). He attempted to relate identification responses to the canonical forms involved; however, he found that correct identification depended more on the quality of the consonant or consonant cluster which separated the vowels. For example, if stops were involved, a much higher correct identification rate was found, but utterances differing in the location of nasals were identified poorly. However, many of his test items differed in syntax; in addition, some utterances involved identifications of portions of words which may also not be appropriate.

McCasland (1974) studied the effects of segmental duration in the perception of juncture. Using phrases like "its still" vs. "it still" vs. "its till" vs. "its dill", he obtained the following results:

1. aspiration of the /t/ determined a parsing response of "its till", and
2. parsing responses for the other three choices were based on a combinatory effect of the duration of the /s/ and the duration of the stop closure.

For the geminate /s/ condition (i.e., "its still"), a long

/s/ was required. In order for the boundary to be heard before the /s/ (i.e., "it still"), the /s/ had to be fairly long, and the /t/ fairly short; for a boundary to be heard after the /s/ (i.e., "its dill"), the /s/ had to be fairly short and the /d/ fairly long. Both segments had to be taken into consideration; subject responses could not be accounted for if only one of the segments was considered.

Repp et. al. (1978) also investigated durational aspects in juncture location. Their study dealt with the differences between 'grey ship' vs. 'great ship' vs. 'grey chip' vs. 'great chip'. They found that both the silence (t-closure) and fricative durations were important in these distinctions. Furthermore, speaking rate affected the relative durations between the silent portion and frication; the noise duration was effectively longer at the faster rate. Results are similar when the 'source' sentence is altered (e.g., using the utterance 'great ship' and altering its silence and frication duration values vs. using the utterance 'grey ship' and altering these segment durations). The authors concluded that since listeners integrate such numerous types of cues into phonetic perception, it is the articulatory act that is perceived. However, there is no clear motivation for their neglect of other models of speech perception.

Perceptual studies of juncture have led to some interesting questions regarding listeners' ability to segment words from an utterance. We know that spectral

allophonic differences play a large role, particularly in differentiating juncture location among simple monosyllabic words. We also know that both stress and rhythm affect listeners' parsing responses in differentiating the syllabic patterns of words in an utterance. Furthermore, we know the general relative importance of spectral, stress and rhythm cues in parsing speech. However, it is not known how utterances involving segements with weak allophonic cues are handled, nor is it discussed how stress and duration patterns affect juncture locations. The following studies address these questions, both from a production and perception point of view.

## III. Measurement Study of Production

This chapter describes a quantitative analysis of intensity and duration differences of segments near juncture points. In addition, the effects of contrastive stress are investigated, as well as effects due to differing types of consonants.

### A. Description and Measurement of Data

#### Speakers

Four speakers, three male and one female, were recorded. All speakers were native Canadians and had no history of speech disabilities.

#### Materials

Four sets of short sentences were constructed. The carrier frame was kept syntactically constant for reasons discussed in Klatt (1975). This frame was: "The sheep like(s) (*verb*). Four verb pairs were chosen such that deleting the initial /s/ of the first verb would result in the second. In addition, it was important that the number of syllables remain constant (cf. results of Lehiste, 1960, regarding elongation of final syllables). These verb pairs were: "sleeping-leaping", "sweeping-weeping", "smashing-mashing", and "snapping-napping". Thus, four sonorant consonants (/m/, /n/, /l/, and /w/) were clustered with /s/, two of which carry no noticeable allophonic spectral differences (namely, /m/ and /n/). Changing the placement of juncture resulted in three distinct meanings

for each set:

1. S#S: The sheep likes sleeping. (single sheep)
2. K#S: The sheep like sleeping. (many sheep)
3. S#C: The sheep likes leaping. (change of verb)

Similarly, other such sets were constructed using the remaining verb pairs. Thus, a total of twelve sentences were constructed (Appendix 1).

In addition, stress was altered contrastively:

The sheep lí̀ke sleeping.

The sheep like sleé̀ping.

Since each sentence was spoken under the two stress conditions, this further doubled the material size to a total of 24 sentences.

**Apparatus**

The instruments below were used in this study. Their technical specifications follow.

1. Microphone: Sennheiser MD 421N, frequency response 30-17000 Hz. ±5 dB; sensitivity .2 mV/microbar at 1000 Hz.; cardioid directionality
2. Tape Recorder: TEAC A-7030, frequency response 50-1500 Hz. ±2 dB; speed 14 ips.; SNR 58 dB.
3. Audio-frequency Filter: Frøkjauer-Jensen type 400, frequency response slope 36 dB/oct.
4. Minicomputer:PDP-12A; word length 12 bits; A/D, D/A convertors 10 bits; operating systems OS/8 and

Alligator.

**Recording**

Subjects were individually recorded in a sound insulated recording room. In order to eliminate possible crosstalk effects, only the left channel of the TEAC was used. In order to regulate the tempo of speaking, each sentence was first presented to the subject from a master tape on which a 'master' speaker was recorded. The master tape was made tempo constant by having the master speaker utter the sentences at the same rate as digitally prepared beats. These beats simulated the rhythm of the sentences. The master tape was presented with a Sony tape recorder over Sony headphones.

Subjects were asked to repeat the sentence at the same rate as the presentation on the master tape, but to do so in a natural manner. However, one speaker (Speaker 3) seemed to speak at a faster rate. The sentences were also written out for the subject in order to prevent slips of the tongue and to make the task easier. Four replications were recorded. In each replication, the subject first recorded the post-juncturally stressed sentences (i.e., stress on *likes*), and then recorded the pre-juncturally stressed sentences (i.e., stress on *sleeping*). The middle two replications were picked for measurement purposes. This procedure diminished warm-up and fatigue effects in speaking.

---

[1]The Alligator programming system, developed by Stevenson and Stephens (1978) is written in OS/8 PAL 12D assembly language and is designed for psychoacoustic experimentation. The system is executable on PDP-12 computers.

## Digital Gating

Digitization was done by an interactive Alligator program. For each stimulus sentence, the portion surrounding the juncture (e.g., for the sentence 'The sheep likes leaping.', the surrounding portion is 'likes leap') was digitized and stored on tape. Only this part of the signal was stored due to the limitations of computer memory. The audio signal from the tape recorder was bandpass filtered (68-6800 Hz.) in order to eliminate 60 Hz. hum and possible speech components above 8 kHz. before digitally gating the signal. Care was taken to avoid signal clipping, while still maintaining the broadest possible range of quantization. The wiring diagram is shown in Fig 1.

## Segmentation and Measurement

Each stimulus was segmented into seven sections via Fortran programming described by Nearey and Hogan (1979). To aid in the segmentation procedure, a spectrum analyzer and a playback device were available. The wiring diagram is shown in Fig 1. The seven sections were defined as follows:

1. 'LAI': from the beginning of the /l/, indicated by the start of periodicity of the waveform, up to the silent portion of the beginning of the /k/-closure
2. K: the silent /k/-closure, indicated by zero amplitude line
3. BUR: indicated by '/k/-burst' spectral peaks ranging in the 1.5-4 kHz. range (Halle et. al., 1957)
4. S: the /s/-noise indicated by random-appearing points in

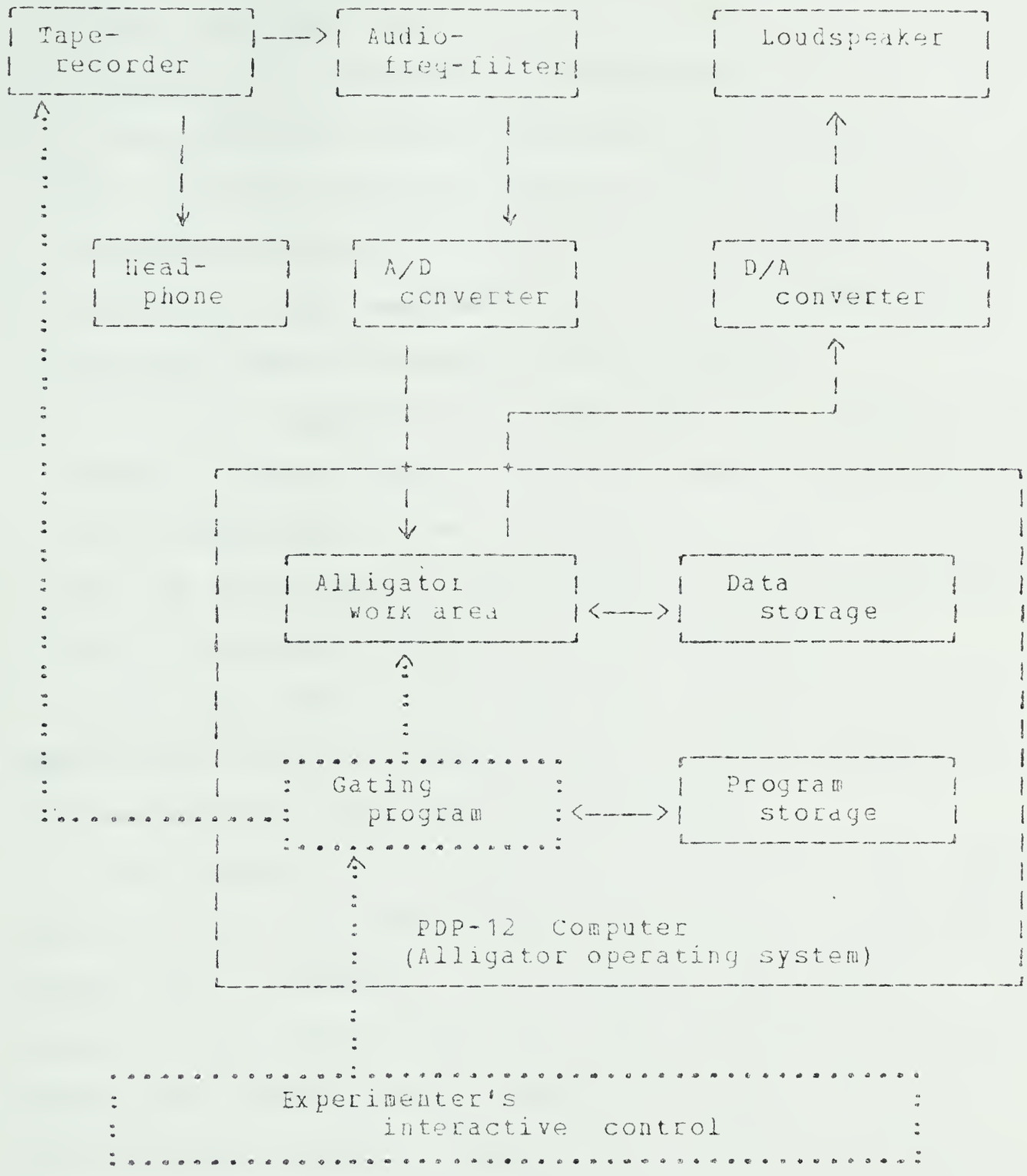



Figure 1. Block diagram of digital gating and segmentation.
*Solid arrows indicate signal flows; dotted arrows, control flows; solid boxes, devices; and doted boxes, controllers.

the waveform and having a spectrum peak at approximately 6 kHz. (Strevens, 1960)

5. P: the pause, indicated by a silent portion
6. C: the consonant from the beginning of periodicity in the waveform following pause or /s/ up to the vowel as defined by formant changes or distinct changes in the pattern of the waveform (as in the case of nasals). Decisions were also based on audio-playback.
7. V: from the beginning of the vowel as defined the formant pattern appropriate to the vowel in question up to the beginning of the silent portion of the following /p/, or in the case of the verb-pair 'smashing-mashing', up to the random waveform of the frication of /s/.

It should be noted that for the /l/ and /w/ cases, segmentation of the consonant from the vowel was difficult, and some arbitrary decisions had to be made.

The duration and intensity of each of the seven sections were then measured by OS/8 Fortran IV programming written by T. Nearey. For intensity measurements, the program first removed the DC bias, and then calculated average amplitude as r.m.s. values. All values were stored in the Amdahl 470/V6 computer for statistical analysis. Measurement means and standard deviations are listed in Appendix 2.

## B. Statistical Analysis: Results and Discussion

The duration and intensity measurements were statistically analyzed using two statistical tests: the analysis of variance (ANOVA) and the discriminant function analysis (DFA). The raw duration data was transformed in two ways, log and square-root conversions, and both ANOVA and DFA were performed for all three measurement scales. Log conversions were investigated because of the report by Port (1978) that all segment durations are roughly proportional to speaking rate. In the log scale, the constant of proportionality becomes an additive term. In this experiment, there were indications of rate variations among speakers, as noted above. Under Port's hypothesis, such variations might be expected to be subsumed by the additive main effects for speakers in the ANOVA. Square-root conversions were investigated, since JND's for noise and tone burst duration are nearly constant for durations in a square-root scale for the range in which most of the present segment durations belong (i.e., 100 msecs.) (Abel, 1971). Intensity measures are reported in decibels.

Considering both ANOVA and DFA results, the square root transformation seemed to be the most appropriate; significant subject interactions decreased, while the proportion accounted by main effects increased in the ANOVA and the Wilks lambda score in the DFA (a measure of group separation) was higher for this transformation.

## Analysis of Variance for Duration

Seven ANOVA's were done, one for each of the seven sections as described by the segmentation procedure. The ANOVA design consisted of four fully crossed factors with two replications in each cell. The four factors were:

1. S: subjects, of which there were four, S1, S2, S3, and S4.
2. A: accent, of which there were two levels; post-junctural stress=A1 and prejunctural stress=A2.
3. C:[2] consonant, of which there were four; /l/=C1, /w/=C2, /m/=C3, and /n/=C4.
4. J: juncture type, of which there were three; S#S=J1, K#S=J2, and S#C=J3.

Table 1 shows the results of the ANOVA's for each section. A conservative F-test (Winer, 1971) was done due to the violation of homogeneity of variance which was found in the duration of the K-closure, and in the amplitude measurements of the following consonant (i.e., /m, n, w, and l/), and the final vowel. The Bartlett Test for homogeneity of variance indicated that these sections showed heteroscidasticity beyond the .01 level. It seemed reasonable to treat all sections in a similar manner; hence, all sections were tested conservatively owing to the lack of homogeneity of variance in some sections. In addition, all reported significant effects are significant to the .01

---

[2] 'C' is meant to designate only the four sonorant consonants in this investigation and does not refer to all consonants, as in the ordinary symbolic convention.

TABLE 1

Anova's for Duration

| Source | Error | D.F. | F Ratio | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 'Lai' | Closure | Burst | Pause | 'S' | Consonant | Vowel |
| S | R(SACJ) | 3 | 93.37** | 14.48** | 20.79** | 2.70 | 7.67** | 26.39** | 20.5** |
| A | SA | 1 | 4.64 | .36 | 3.87 | 4.43 | 1.56 | 2.71 | 9.06 |
| C | SC | 3 | 1.38 | .79 | .21 | 1.86 | 5.28 | 1.13 | 77.55** |
| J | SJ | 2 | 3.32 | 44.48** | 5.84 | 33.43** | 52.71** | 107.88** | 14.2** |
| SA | R(SACJ) | 3 | 12.37** | 6.03* | .90 | .25 | 3.21 | 5.33* | 3.01 |
| SC | R(SACJ) | 9 | 1.16 | 1.44 | 1.27 | 5.13* | 1.9 | 6.17* | 5.56* |
| AC | SAC | 3 | 1.46 | 1.59 | 2.66 | 1.79 | 3.23 | .69 | .47 |
| SJ | R(SACJ) | 6 | .71 | 1.36 | 2.01 | 2.7 | 8.03** | 1.61 | 2.8 |
| AJ | SAJ | 2 | 1.87 | .12 | 3.38 | 4.43 | 2.34 | 1.71 | 2.75 |
| CJ | SCJ | 6 | 4.92* | 1.36 | 1.17 | 1.86 | 1.71 | 1.61 | 1.43 |
| SAC | R(SACJ) | 9 | .96 | 1.85 | .24 | .38 | .23 | 1.56 | 1.42 |
| SAJ | R(SACJ) | 6 | .67 | 1.63 | .51 | .25 | 1.6 | 2.93 | .61 |
| SCJ | R(SACJ) | 18 | .13 | .88 | .88 | 5.13* | 1.56 | 1.9 | 1.37 |
| ACJ | SACJ | 6 | .75 | 1.09 | .40 | 1.79 | 1.27 | .29 | .90 |
| SACJ | R(SACJ | 18 | .51 | 1.25 | 1.11 | .37 | .61 | .81 | 1.73 |
| R(SACJ | | 96 | | | | | | | |

level.

For all sections except the pause section, speaker main effects were significant. Significant main effects due to juncture were evident for all sections except for the LAI and burst portions. In other words, the different types of junctures affected the durations of five out of the seven sections measured. Significant consonant main effects appeared in the analysis of the vowel section only. Thus, the duration of the vowel was affected by the type of consonant preceeding it. Two significant interactions emerged. The LAI portion showed a Subject-Accent (SA) interaction, indicating a speaker dependent effect of stress on the vowel. The S section showed a significant Subject-Juncture (SJ) interaction. Fig. 2 shows that the main effects of juncture (described below) are maintained.

**Tukey Tests for Duration**

Tukey tests, designed to test for significant differences between levels within each factor were done for all significant main juncture effects. Main effects due to subjects were not analyzed since it is the effects of juncture, and not speaker differences which are of interest. Table 2 summarizes the results of these tests.

The durations of the K in the K#S condition are significantly greater than K durations in either S#S or S#N. That is, a non-clustered K at the end of the word (like) is significantly longer than a clustered K (likes).

S durations in S#S were significantly greater than S

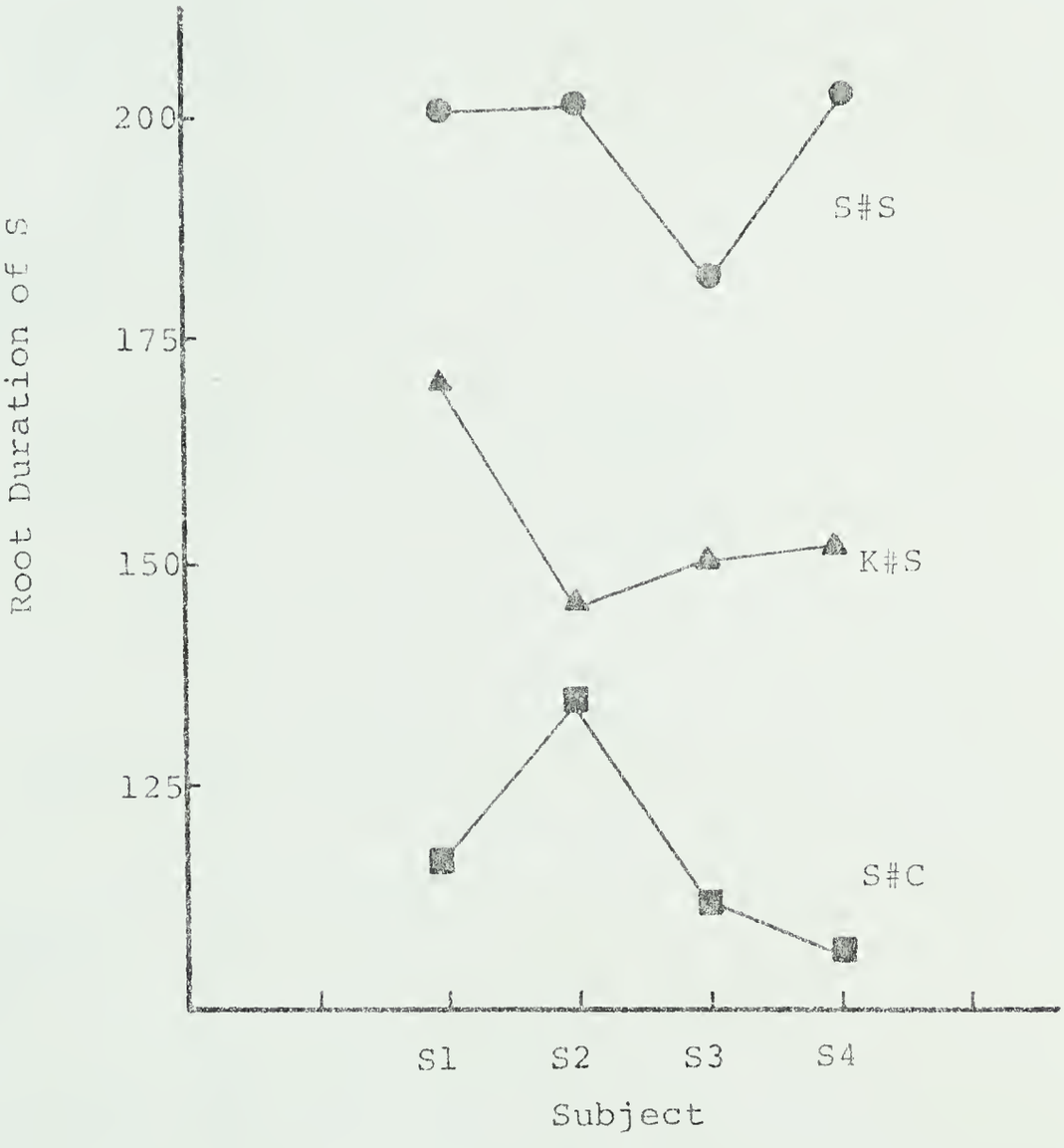



Fig 2. Subject-Juncture Interaction

TABLE 2

Tukey Tests for Duration

| | Means | | | F Ratio | | | Summary |
|---|---|---|---|---|---|---|---|
| | J1 | J2 | J3 | J1-J2 | J2-J3 | J1-J3 | |
| K Closure: | 6.53 | 7.98 | 5.97 | 9.31** | 12.93** | 3.63 | J2 > J1 = J3 |
| S : | 14.01 | 12.41 | 10.8 | 7.23** | 7.23** | 14.53** | J1>J2>J3 |
| Consonant: | 8.37 | 6.49 | 6.46 | .30 | 18.15** | 17.85** | J3 >J2 = J1 |
| Vowel : | 11.53 | 10.82 | 10.77 | .50 | 6.76** | 6.26** | J3>J2 = J1 |

KEY: J1 = S#S

J2 = K#S

J3 = S#C

durations in K#S, and those in K#S were greater than those in S#C. That is, a double /s/ is longer than a single /s/, and furthermore, a word-initial /s/ is longer than a word-final /s/. The consonant following the /s/ is longer in S#C than in K#S or S#S, which means that an initial non-clustered sonorant consonant is longer than an a non-initial, clustered consonant. Finally, the vowel portion is longer in S#C than in K#S or S#S. That is, vowels following single consonants are longer than vowels following clustered consonants. However, this last result is based on a somewhat arbitrary segmentation procedure, as it was difficult to find the exact location of where the consonant ended and the vowel began. The differences in pause were not analyzed since the pause was found only in the S#C condition.

For the significant main effects of consonant on the vowel section, a Tukey test was also done. It was found that vowels following nasals were longer than vowels following liquids. [3] However, due to the arbitrary cut-off decisions while measuring liquid consonants, this result is difficult to interpret.

**Analysis of Variance for Intensity**

Five ANOVA's were done, one for each section except "pause" and "K closure". (These sections were omitted since

---

[3]Tukey tests showed that the following differences were significant: /n/ vs. /w/, $F(4,9)=5.99$; /m/ vs. /w/, $F(3,9)=5.82$; /l/ vs. /n/, $F(3,9)=5.96$; /l/ vs. /m/, $F(2,9)=5.79$. The differences between /l/ vs. /w/ and /m/ vs. /n/ were not significant.

they are normally segments of zero intensity.) The ANOVA design was identical to the analysis of duration and thus had the four fully crossed factors of Subject, Accent, Consonant and Juncture type. Table 3 shows the ANOVA results for the five sections. For all sections, speaker main effects were significant, indicating either large speaker variability and/or possible minor effects in the recording situation. Juncture affected the /s/ and consonant portions but only to a .05 significance level; this is interesting because only these sections are in proximity to the boundary location. All sections except the K-burst showed Subject-Accent interactions. It is evident that intensity of elements is affected by contrastive stress; elements in a stressed word are generally more intense than elements which are not in a stressed word. However, subject differences do exist, as evidenced by the SA interaction.

**Tukey Tests for Intensity**

Tukey tests were done only for juncture effects, as subject differences are not under discussion. Results are shown in Table 4.

S's in S#C are less intense than S's in S#S or K#S. That is, final S's are less intense than S's in the other conditions. Similarly,the consonant portion in S#C was less intense than in S#S or K#S. Thus, initial sonorant consonants are less intense than consonants within words. It is important to note that these intensity measures deal with overall intensity values and not intensity shapes. (For

TABLE 3

Anova's for Intensity

| Source | Error | D.F. | F Ratio | | | | |
|---|---|---|---|---|---|---|---|
| | | | 'Lai' | Burst | 'S' | Consonant | Vowel |
| S | R(SACJ) | 3 | 86.67** | 31.82** | 52.31** | 10.29** | 15.61** |
| A | SA | 1 | 1.78 | .88 | .4 | 4.53 | 6.93 |
| C | SC | 3 | 4.7 | 6.08* | 4.03 | 2.96 | 3.62 |
| J | SJ | 2 | .84 | .78 | 9.65* | 6.36* | .71 |
| SA | R(SACJ) | 3 | 10.91** | 1.08 | 29.21** | 27.82** | 66.03** |
| SC | R(SACJ) | 9 | 2.22 | .78 | 2.39 | 4.10* | 1.73 |
| AC | SAC | 3 | 8.23* | 1.01 | 1.25 | 2.89 | 5.43* |
| SJ | R(SACJ) | 6 | 1.51 | .71 | 2.07 | 2.6 | 1.25 |
| AJ | SAJ | 2 | 6.01* | .49 | .45 | 2.32 | .06 |
| CJ | SCJ | 6 | .57 | .06 | .64 | 5.08* | 1.95 |
| SAC | R(SACJ) | 9 | .48 | .22 | 2.17* | .48 | .34 |
| SAJ | R(SACJ) | 6 | .08 | .48 | .76 | .38 | .12 |
| SCJ | R(SACJ) | 18 | .81 | .78 | .81 | .49 | .42 |
| ACJ | SACJ | 6 | 1.26 | .96 | .55 | .85 | 1.25 |
| SACJ | R(SACJ) | 18 | .37 | .56 | .83 | .27 | .26 |
| R(SACJ) | | 96 | | | | | |

TABLE 4

Tukey Tests for Intensity

| | Means | | | F Ratios | | | Summary |
|---|---|---|---|---|---|---|---|
| | J1 | J2 | J3 | J1-J3 | J2-J3 | J1-J2 | |
| S: | 22.08 | 21.62 | 19.98 | 5.92* | 5.60* | 1.30 | J3 < J1 = J2 |
| Consonant: | 36.21 | 36.39 | 34.24 | 4.164 | 4.54 | .38 | J3 > J2, J2 = J1 |

KEY: J1 = S#S

J2 = K#S

J3 = S#C

example Lehiste, 1960, found that the intensity of initial /n/'s rose slower than non-initial /n/'s.)

**The Discriminant Function Analysis**

A linear Discriminant Function Analysis (DFA) was done on all duration and intensity measurements combined. This analysis classifies data into groups according to functions based on variables. For this experiment, the functions were based on the duration and intensity of the seven measured sections. The groups into which the data was classified were the three juncture categories, where S#S=Category 1, K#S=Category 2, and S#C=Category 3.

Table 5 shows the classification results. Overall, the groups were correctly classified 91.67% of the time, based on all duration and intensity measurements. The group which was classified correctly to the highest extent was S#C. S#S was correctly classified the second highest, and finally, K#S was classified the poorest. Fig. 3 shows the scatterplot diagram, where * indicates the group centroid. The canonical discriminant function 1 is most strongly correlated with the S-duration, duration of the consonant, information about the pause, and to a lesser extent, on the S-intensity. In general, however, S-intensity and S-duration are correlated. Function 2 is most strongly correlated with the K-duration and S-duration. As the scatterplot shows, Function 1 distinguishes S#C from K#S and S#S, while Function 2 distinguishes K#S from S#S. Thus, duration of the S, C, and P serve to separate S#C from the other two juncture

TABLE 5

Classification Results of Discriminant Function Analysis

| Actual Group | No. of Cases | Predicted Group Membership | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Group 1 (S#S) | 64 | 59<br>92.2% | 5<br>7.8% | 0<br>0.0% |
| Group 2 (K#S) | 64 | 9<br>14.1% | 55<br>85.9% | 0<br>0.0% |
| Group 3 (S#C) | 64 | 0<br>0.0% | 2<br>3.1% | 62<br>96.9% |

Fig 3. Scatterplot of Discriminant Function Analysis

conditions, while the double and single S conditions are separated on the basis of the S and K durations. In addition, the plot shows that S#C is separted from the other conditions to a larger extent, but S#S and K#S often overlap. (This is also reflected in the classification scores above.) Apparently, subjects produce a greater difference between "likes napping" and the other two sentences than between "likes snapping" vs. "like snapping".

**Summary**

This measurement study showed that the three categories of juncture are differentiated in speech production, and further pointed to those elements within the signal that are affected as a result of juncture location. Both the ANOVA and DFA analyses pointed to the duration of the S as being the most highly affected element as juncture is varied. Other elements which are of importance are duration of the consonant, K-closure, and pause.

In general, S#C produces a short K, a short non-intense S, a pause, and a long N. This reflects the cluster at the left side of the juncture, and the single consonant at the right side. K#S produces a long K closure, an S with medium duration and high intensity values, no pause and a short N. This reflects the single consonant at the left side of the juncture, and the cluster following the juncture. S#S produces a long S of medium intensity, a short K closure, no pause and a short N. This reflects the pre-junctural cluster and the conjoining of the two /s/'s in production.

## IV. Identification Tasks

In the previous chapter, several aspects of the signal differed in duration according to juncture location, namely the durations of K, S, C, and P. This chapter addresses whether these changes alter subject's perception of juncture. Two perceptual experiments were done. They were designed to establish sets of stimuli that were largely identified as representing the three juncture types. These experiments were done in order to test whether alterations observed in the signal (as defined by production measurements) cued word boundary location for listeners. The first experiment used a forced choice method, whereas the second used a free response mode. It was of interest to see whether the response task altered subjects' identifications. It was also of interest to see whether 'typical' or average duration values for each juncture category would indeed signal that category.

Due to computer storage restrictions, as well as the desire to limit the number of stimulus items, only one speaker, one stress condition, and one verb-pair were used. The utterances of Speaker 2 were chosen due to their consistency and low misidentification rate in the DFA. Also, average values for this speaker coincided to a high degree to the averages pooled across subjects. The 'normal' post-junctural stress mode was employed (i.e., likes napping) due to the absence of accent differences (in relation to juncture) in the ANOVA tests. Finally, the verb

pair "snapping-napping" was chosen because the segmentation problems were considerably less than with the liquid consonants /l/ and /w/; the "smashing-mashing" pair could have been utilized as well, but it was a bit more difficult to segment a following /s/ than a following p-closure.

## A. Experiment 1: Identification by Forced Choice Method

This first experiment used the method of forced choice in order to obtain identification curves of juncture location.

### Preparation of Stimuli

The stored portions of Subject 2's "likes snapping" and "likes napping" were used in this experiment. In addition, the subject's beginning and final portion of an originally taped sentence was sampled, digitized and stored in the same manner as described above. From these utterances, a total of 31 segments was made. There was one beginning segment SHLK ("The sheep lai"), ten K-closures of varying durations, six pause (P) lengths, five different durations of N, eight S durations and one end portion "APPING". For each of K, S, N and P, there were three segments corresponding to typical or average values in each juncture category and ambiguous portions consisting of mean values between each typical value. The segments were prepared using interactive Alligator programming and their detailed characteristics are listed in Appendix 3. Their preparation is described below.

For N's, the long nasal portion of the originally taped

"napping" portion was segmented. Beginning and end portions were segmented out, using zero-crossings as end-points, and stored. A single pitch period from the middle portion of this long N was subsequently segmented at zero-crossing end-points, and stored. New N's were made by digitally queuing the beginning portion, with a number of middle portions, and the end portion. By varying the number of middle portions added, the duration of the segment could be varied. Two 'typical' N's were constructed, one for S#C, and one for S#S and K#S. The latter two juncture conditions required only one representative as the measurement study revealed no significant difference between N durations in these categories. A short N, a long N, and one 'ambiguous' N (having a duration between S#C and K#S/S#S) were also constructed.

For the S's, 50 msecs. of the beginning and end portions of the subject's S in "likes snapping" were segmented out and stored. The burst of the /k/ was included in the beginning portion due to possible segmentation problems. The middle portion of this fairly steady-state /s/ was subsequently segmented at varying degrees from the centre, providing several different durational values. New segments were made by queuing the beginning portion, any one of the several middle portions, and the end portion. Three 'typical' and three 'ambiguous' S's were made, as well as a short S, and a long S. In addition, the intensity values of each 'typical' S were altered so as to coincide with average

intensity values for each category. The intensities of the 'ambiguous' S's were altered to be the mean average between intensity values of two categories. Like the durational values, intensity values for each category were based on the previous measurement study. The short and long S intensities were altered to typical S#C and typical S#S intensity values, respectively.

The P (pause) and K-closures were made by using small silent portions and stringing any number of these portions together in order to produce varying durational levels of silence. The original portion was from a pause in the speaker's S#C category.

The duration and intensity values of all segments are listed in Appendix 3.

**Lists of Stimuli**

For this experiment, it was of interest to test the perceptual categorization of 'typical' and 'ambiguous' frames. Three of each of such frames were constructed and presented using Alligator programming. The segments for each of these frames were:

1. Frame 1: 'Typical S#S'; SHLK (=SHEEP LIKE), typical K in S#S, long S, no pause, typical N in S#S/K#S, APPING.
2. Frame 2: 'Typical K#S'; SHLK, typical K and S in K#S, no pause, typical N in S#S/K#S, APPING.
3. Frame 3: 'Typical S#N'; SHLK, typical K, S, P and N in S#N, APPING.
4. Frame 4: 'Ambiguous between S#S and K#S'; SHLK,

ambiguous K, S, and N (mean values between) S#S and K#S, no pause, APPING.

5. Frame 5: 'Ambiguous between S#S and S#N'; SHLK, ambiguous K, S, P, and N between S#S and S#N.
6. Frame 6: 'Ambiguous between K#S and S#N'; SHLK, ambiguous K, S, P, and N between K#S and S#N.

A random list of six of each of these frames was produced, providing 36 stimuli. All stimuli were repeated twice. This set of stimuli was desampled by passing it through the FroKjauer-Jensen filter described in Chapter 3, and recorded on audio tape.

**Subjects**

Eight linguistically-trained subjects, with no known hearing difficulty, participated in this experiment. Seven of the listeners were native speakers of Canadian English, while the eighth, a trained phonetician, was a native speaker of American English. Only the latter was familiar with the details of the experiment. In addition, three subjects did two experiments, while the remaining five did the test only once. Since this was a pilot test, the experiment was treated as if there were eleven subjects, each doing the test only once.

**Apparatus**

The tape-recorded stimuli were presented to subjects. The instruments used for this presentation were as follows:

1. Power Amplifier: Braun AG Type CSV 250
2. Tape Recorder: TEAC A-7030

3. Headphone Sets: Telephonics TDH-49, frequency response 30 to 6000 Hz ±3 dB.

The wiring diagram is shown in Fig 4.

**Procedure**

Listeners were given a blank response sheet and a 'Key' sheet on which was written the three juncture conditions and their corresponding response number. They were instructed to write a 1 on the answer sheet if they heard "The sheep likes snapping", a 2 if they heard "The sheep like snapping" and a 3 if they heard "The sheep likes napping". Subjects were asked to make their decision after the second repetition of each stimulus; this was done in order to reduce possible auditory contrast effects. They were also given a 'Key' sheet on which was written the three juncture conditions and their corresponding response numbers. Subjects were requested to refer freely to the key sheet, so as not to forget the juncture-number correspondence.

**Results and Discussion**

The listeners' responses were tallied for each frame. The totals are listed in Table 6. As can be seen by these totals, a S#S/S#N ambiguous frame was no different from a K#S/S#N ambiguous frame, so the latter was subsequently removed from the analysis. The identification curves which resulted are shown in Fig 5. These curves show that average values for S#S, K#S, and S#N consistently elicited the highest responses from listeners in those categories. The ambiguous frames were crossover, or boundary points.

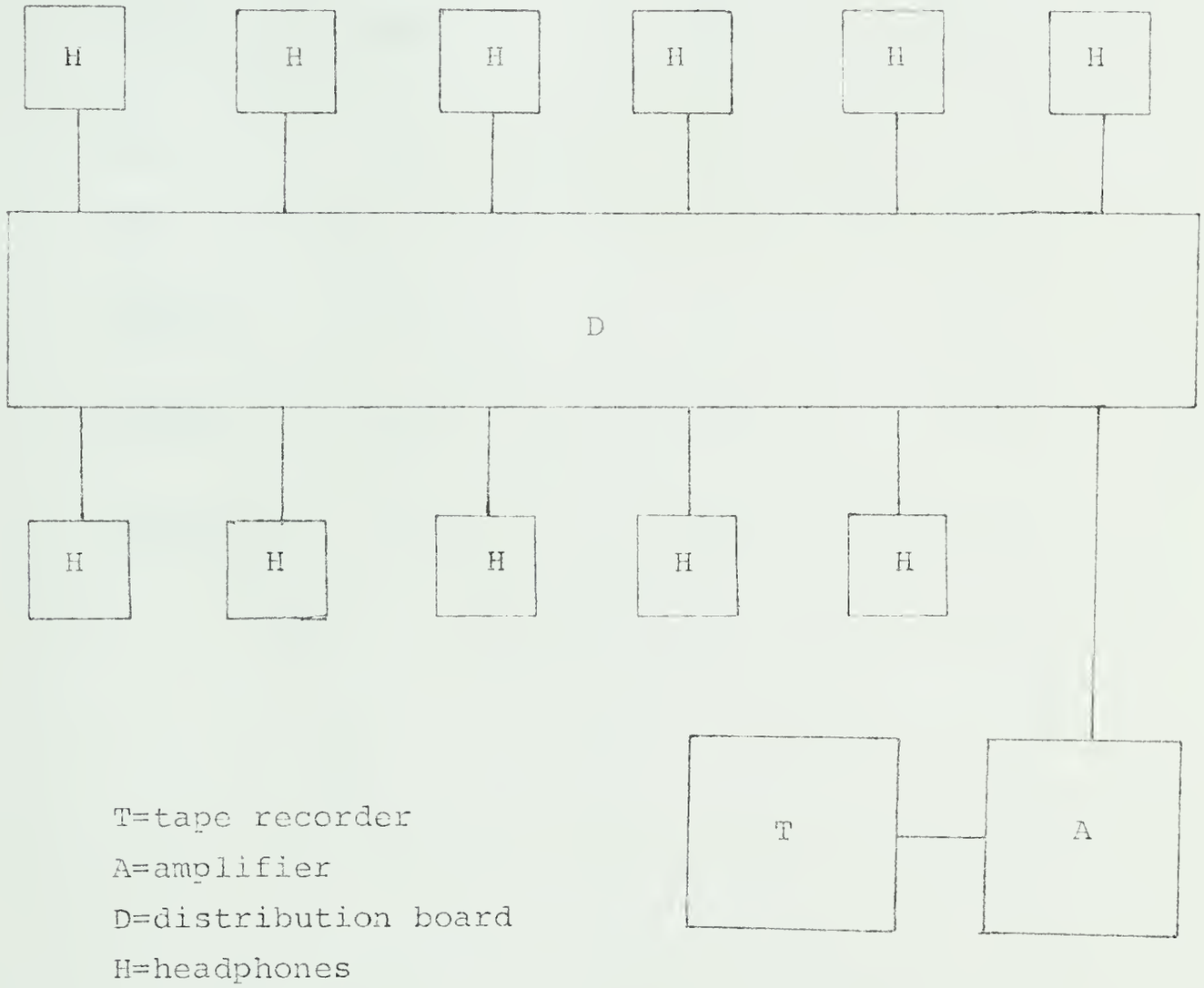



Fig 4. Block Diagram for Signal Presentation

TABLE 6

Subjects' Responses for the Forced Choice Experiment

| Frame Stimulus | Responses 1 S#S | 2 K#S | 3 S#N |
|---|---|---|---|
| S#S | 34 | 21 | 0 |
| K#S | 21 | 34 | 0 |
| S#N | 6 | 6 | 43 |
| Ambiguous S#S/K#S | 28 | 27 | 0 |
| Ambiguous S#S/S#N | 15 | 32 | 8 |
| Ambiguous K#S/S#N | 15 | 30 | 10 |

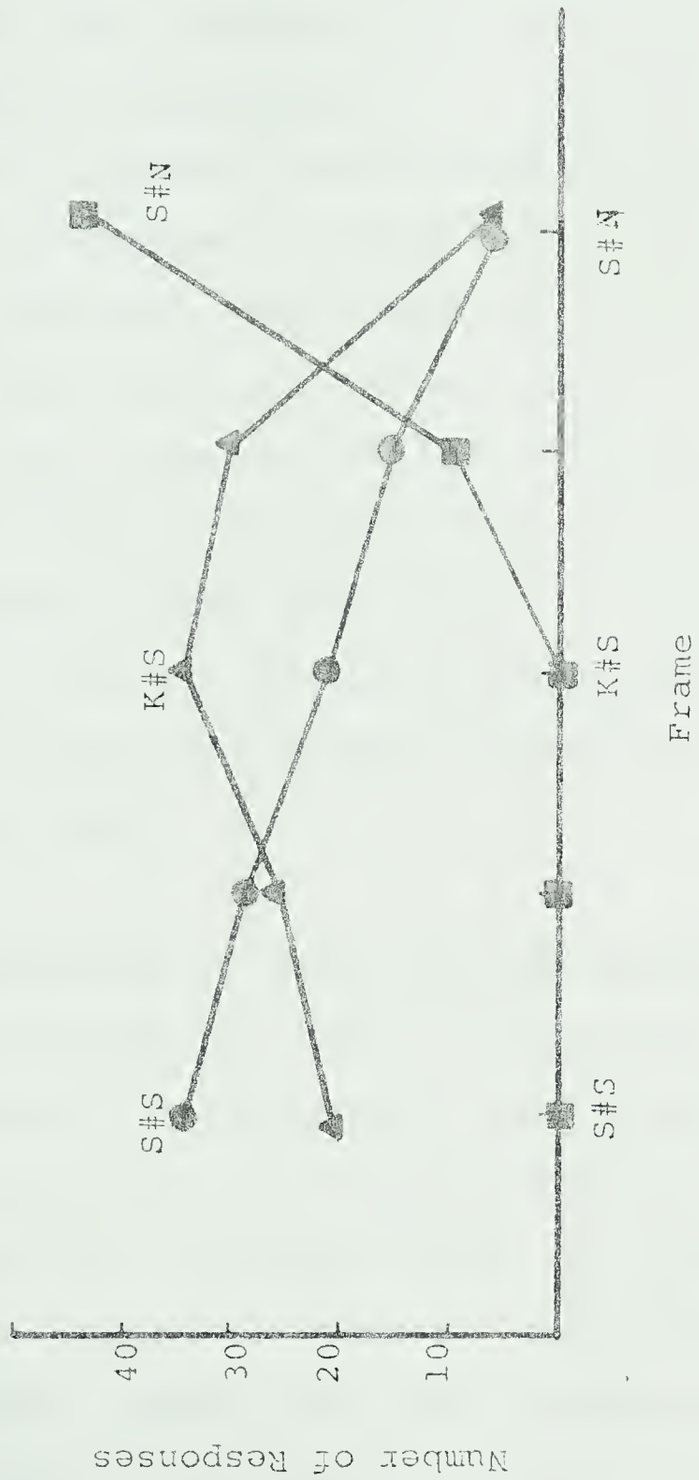



Fig 5. Identification Curves for the Forced Choice Experiment

However, S#N had a much larger percent of 'correct' identifications than either of the other two categories. Also, the responses for the 'ambiguous S#S/K#S' frame were more evenly split than in the 'ambiguous K#S/S#N' case. A possibility of potential difficulties with a forced choice procedure became apparent; it was possible that in a free choice situation, the S#S response might not even appear. In addition, subjects reported that a distinction between S#S and K#S was often very difficult to make, whereas S#N sentences were easily identifiable. A response preference to the K#S category in the 'difficult' cases was suspected, since there were larger totals in this category for all ambiguous frames (see Table 6). Perhaps this preference was due to the fact that the K#S response number (2), was always positioned in the middle of the two alternate choices on the answer sheet.

## B. Experiment 2: Identification by Free Choice Method

In order to ascertain the extent of any response bias, if in fact there was one, a second experiment was conducted. This experiment replicated Experiment 1 with one variation. Choices of responses were not given to the listeners; rather, subjects wrote down what they heard. It was of interest whether the S#S category would emerge under this condition. If no subjects listed this as an alternative, it would be an indication that the previous identification curves were a result of the forced choice method used.

## Subjects

There were eight linguistically-trained subjects, none of whom had participated in Experiment 1. All subjects were native speakers of Canadian English, and had no known hearing disabilities.

## Stimuli and Apparatus

The stimuli and apparatus were identical to those used in Experiment 1.

## Procedure

Subjects were instructed to listen to the presented sentences and write down what they heard. Sufficient time was provided for these written responses.

## Results and Discussion

Subjects's responses were totalled and identification curves generated (Fig 6). Subject differences emerged. Two subjects had no S#S category at all; they did not hear any of the sentence frames as having a double /s/. Of these two listeners, one responded with K#S responses to all presented stimuli. For all subjects, again, more responses were listed as being in K#S than in any other juncture category. This suggests that there are individual differences regarding the extent of a K#S preference. Also, the fact that, for some subjects, the S#S category was not there indicates that either it is harder to signal this category, or the subjects did not 'think of' all the possible alternatives. There is some indication that the latter explanation was at work due to the observation of one subject's response behaviour. This

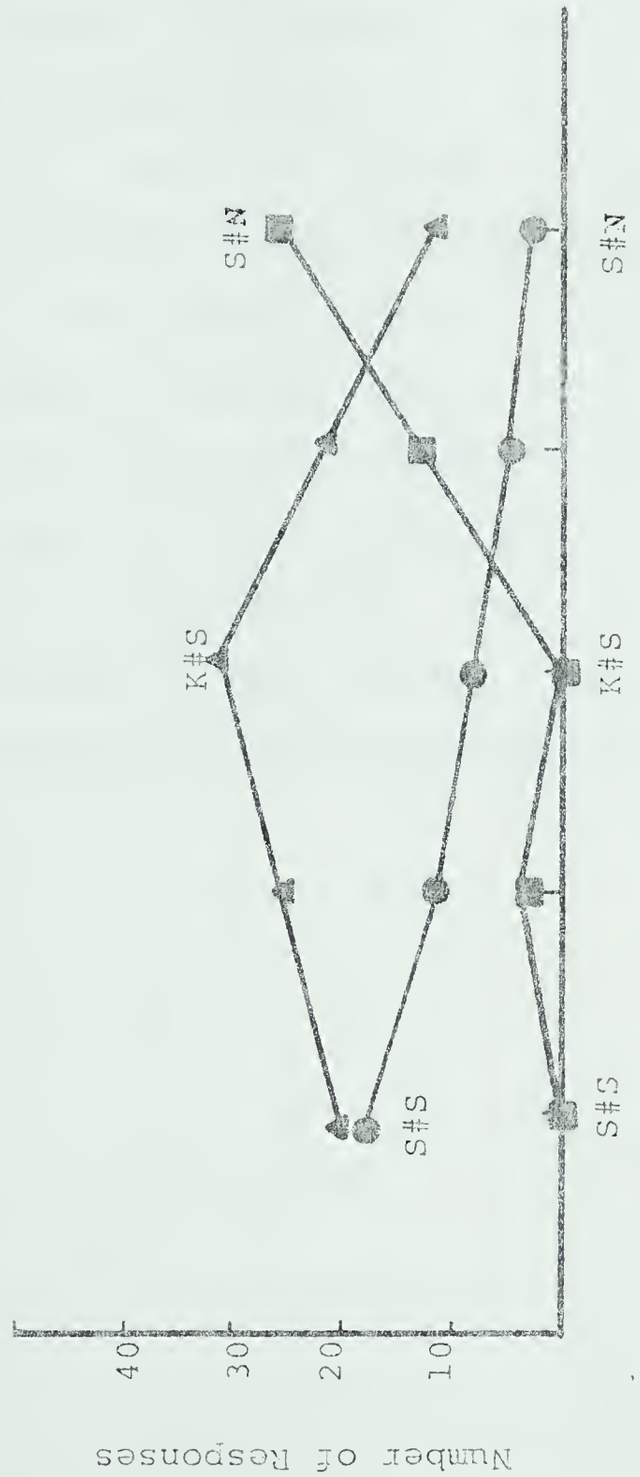



Fig 6. Identification Curves for the Free Choice Experiment

subject, up to the middle of the experiment, did not respond with any S#S's, but after the 'discovery' of this alternative choice, began responding in the same pattern as those who had all three categories from the beginning.

Indeed, the majority of subjects responded with all three categories. When only their identification curves are plotted (Fig 7) these curves are similar to the identification curves obtained by the forced choice method. Thus, for subjects with all three categories, either method of response produces similar results.

This experiment therefore indicates that while there are some listener differences, the majority of subjects respond similarly in the two response conditions. Furthermore, the forced choice method offers the experimenter a controlled response structure.

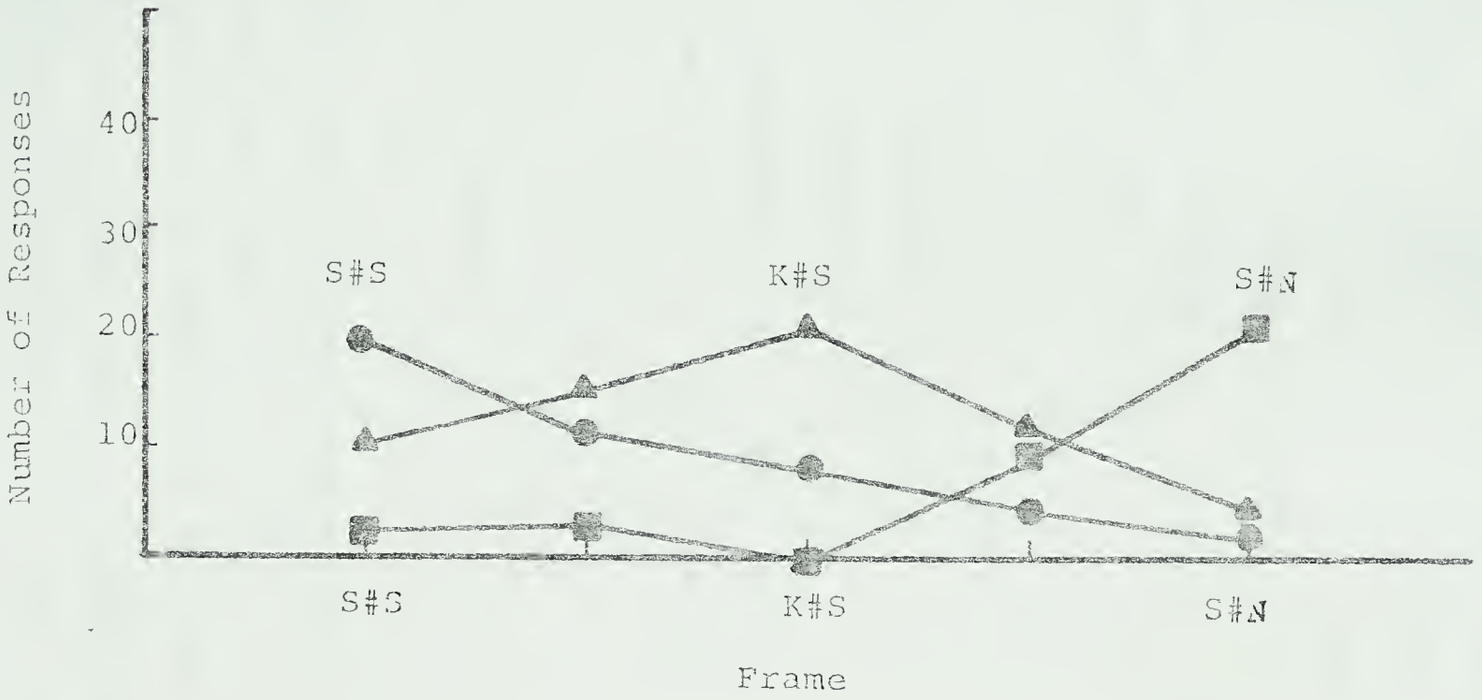



Fig 7. Identification Curves for the Free Choice Experiment for subjects responding with all three categories

## V. Combinatory Perceptual Studies of Juncture

From the previous pilot tests, it is known that changes in the signal alter listener's perception of word boundary location. This chapter addresses what particular aspects of the signal account for such changes. Furthermore, it is of interest to know how various combinations of cues are handled by listeners, particularly when these 'weaker' durational allophonic changes are involved. Thus, experiments were designed in order to test the combinatory effects of the K, S, P and N elements.

In designing such combinatory experiments there are numerous problems associated with the statisical design. There is a general lack of statistical experience throughout the literature, and proper statistical models for such experiments have not been developed. Nonetheless, it is important to explore the effects of a multitude of factors, and perhaps such experimentation will spur the development of stastical methods that can better handle these problems.

It was of particular interest to find out what listeners would do when segment durations conflicted. What would subjects hear when elements were artificially combined that were not in combination in actual production? Is there a pattern to these types of judgements, or are they random? Do some specific elements of the signal 'override' other cues? Finally, do the major cues K, S, P, and N act independently in some sense, or do they interact? These experiments, therefore, test the limits to which

combinations of elements can be classified, and which specific elements are most important in the decision making process of the listener.

## A. Experiment Three: Swamping Experiment

This experiment was exploratory in nature, to see if a change in only one segment of a frame would alter subjects' responses. Would any such elements completely override the cues in the frame? Would such cues 'swamp' listeners' judgements to one category only? This pilot test was performed in order to test for levels of segments that would make a difference in perception, but not override subjects' judgements completely. It was of concern to limit the duration of the elements to those found in production. It was also of concern to define 'perceptual' ranges, so that in the fully crossed experiment, all elements would show a meaningful cross-over point in categorization.

In addition, a type of rating task was simultaneously administered in order to test subject's strength of responses. This was done in order to obtain a rough estimate of how subjects classified utterances which seemed somehow odd to them for the category chosen.

### Subjects

Ten linguistically trained subjects, with no known hearing difficulty, participated in this study.

### Lists of Materials

The three typical frames used in Experiment 2 were

again used. Four extreme levels of each of K, S, P, and N were picked from the extreme ranges found in production. Two of these levels were short while two levels were long. These levels of elements were substituted one at a time into the frames; the other remaining elements in the frame were held constant. Thus, only one substitution was done for any one frame at a time. All levels were within the limits of observed production measurements. The four levels of each segment were:

1. K: typical K in S#N, ambiguous K between S#S and S#N, a fairly long K, and a very long K.
2. S: a short S, a typical S in S#N, a typical S in S#S and a long S.
3. P: no pause, a short P, typical P in S#N, and a long P.
4. N: a short N, ambiguous N between K#S and S#N, a typical N in S#N, and a long N.

(See Appendix 3 for actual duration and intensity values.) This yielded a total of 48 stimuli (i.e., four levels of four variables in three frames). Two randomizations were done thereby increasing the number of stimuli to 96. A digitally prepared sine tone was inserted after every five sets of repeated stimuli in order to allow subjects to keep their proper location on the answer sheet. The stimuli were recorded on audio tape.

**Apparatus**

The same instrumentation was used as in Experiment 1.

## Procedure

The identification task was identical to that explained in Experiment 1. Subjects were asked to record, on computer-scoreable answer sheets, a 1, 2, or 3, depending on their judgements of S#S, K#S, or S#N categories. In addition, subjects were asked to tick off the number of the question if the utterance seemed, in their judgement, to be a particularly odd representative of that category. The 'Key' sheet was again provided for their convenience.

## Results and Discussion

Four identification curves for each of the three frames were generated, one for each variable (K, S, P and N). These curves are shown in Figs 8. Each variable will be discussed separately.

In some cases, the effects of the elements with respect to one another is discussed; it is important to note that the different elements cannot really be equated, since it is not known whether a particular duration level of the S, for example, is comparable to that same duration of the K (or any other element). Therefore, it must be kept in mind that these elements may have different psychophysical effects from one another. Another point of caution is that the range of duration is different for each element: the S ranges from 100-200 msecs., the K from 23-85 msecs., the P from 0-72 msecs., and the N ranges from 40-80 msecs. Thus, the ratios between the longest and shortest elements differ between these four elements.

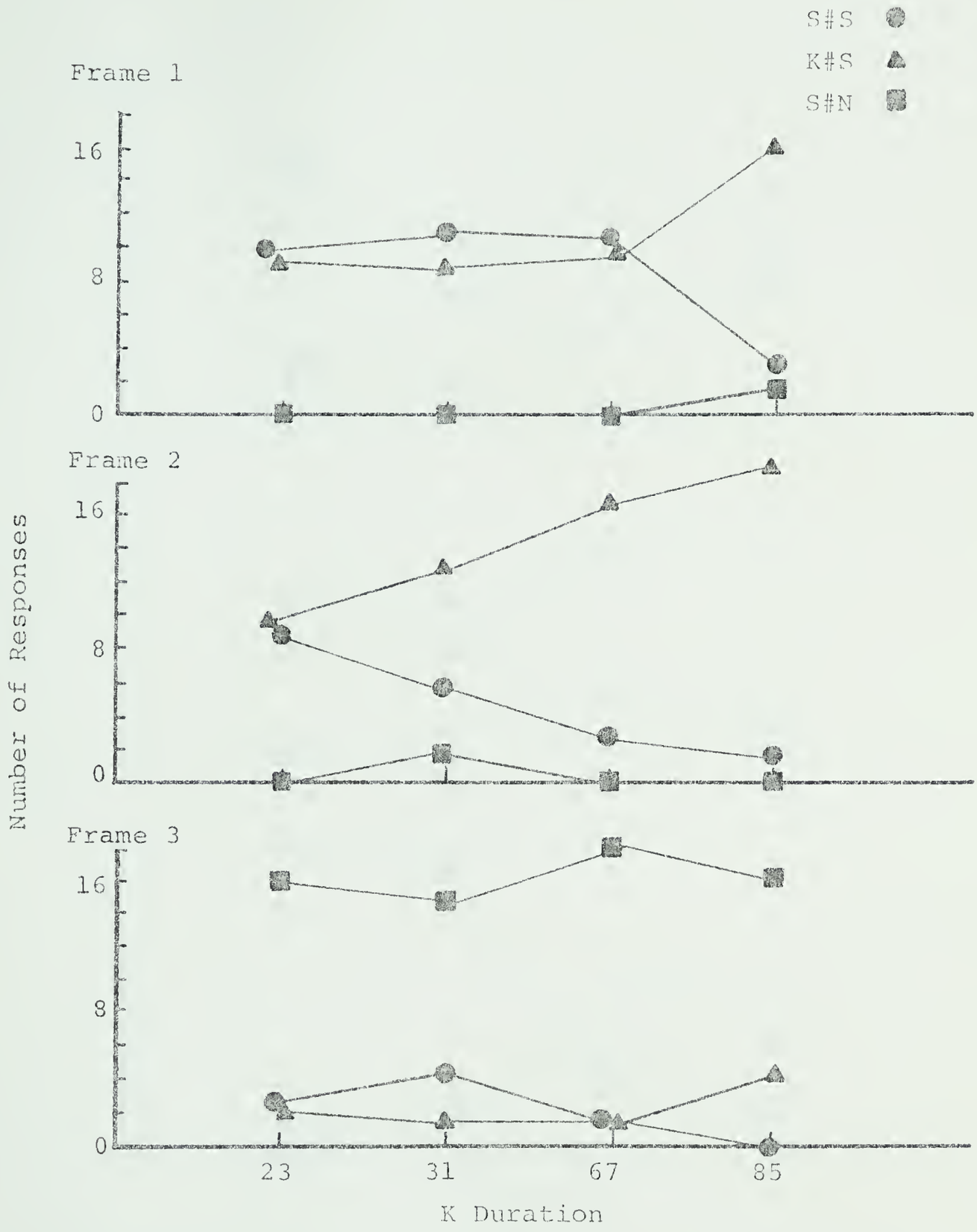



Fig 8. Swamping Curves

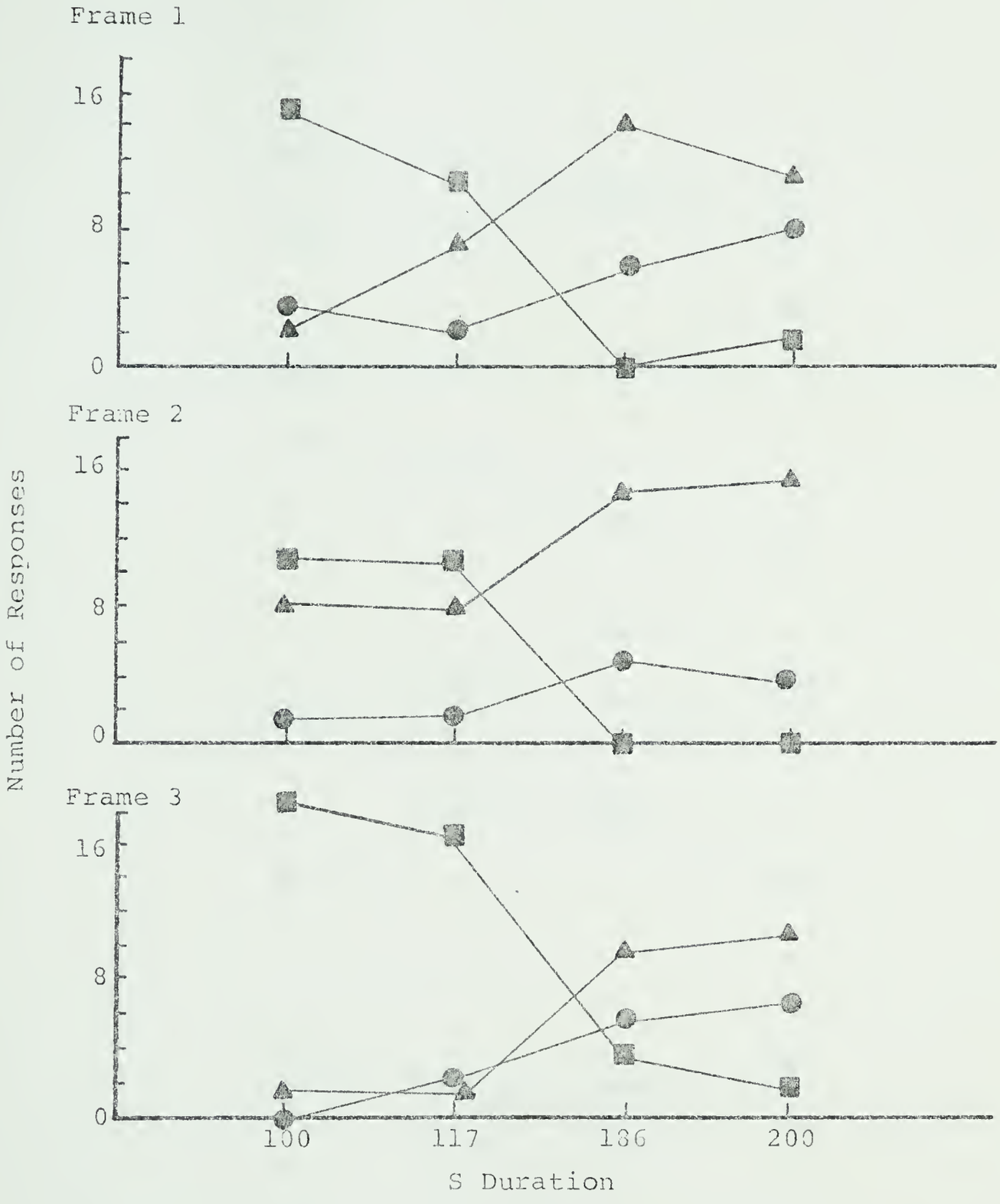



Fig 8. Swamping Curves
(continued)

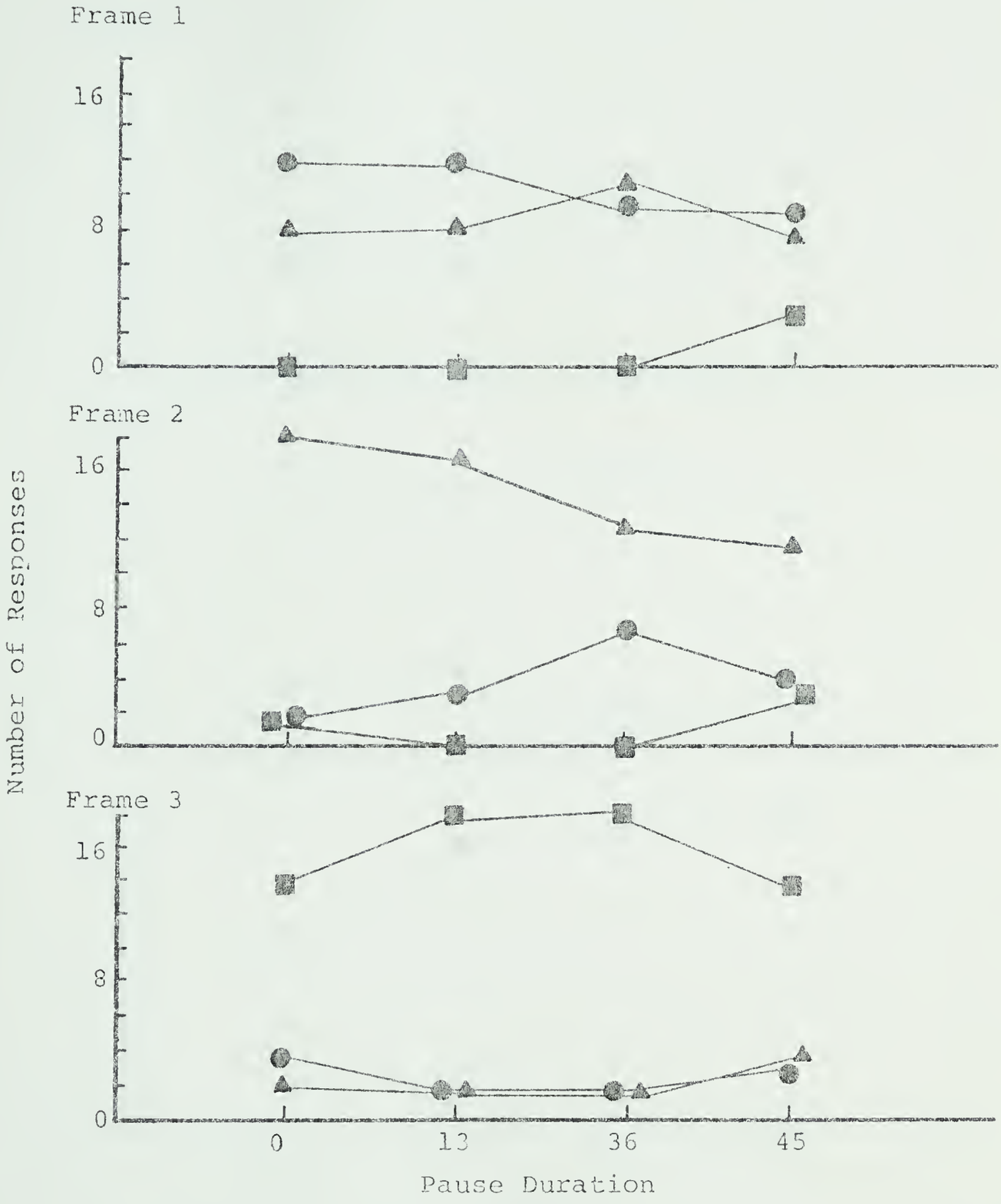



Fig 8. Swamping Curves
(continued)

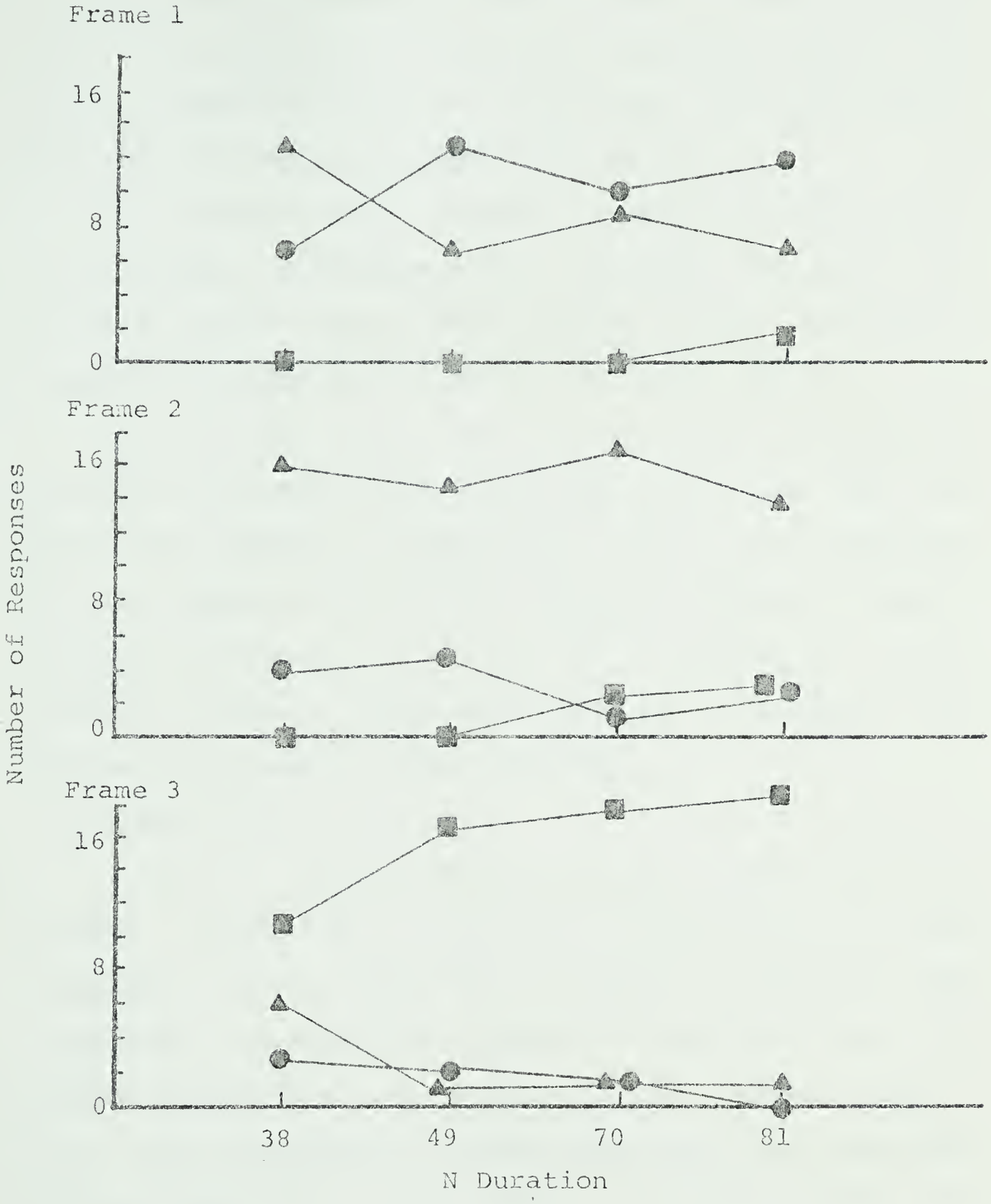



Fig 8. Swamping Curves
(continued)

The Effects of K

For the S#S frame, a long K would 'override' the long S cue and more K#S responses resulted. Shorter K's in this frame are typical of the frame, and therefore, subjects responded with more S#S responses.

For the K#S frame, longer K's, which are typical of this category, yielded more K#S responses. Shorter K's elicited more S#S or S#N responses, since the short K signals a cluster at the end of the word "likes".

For the S#N frame, the K was not powerful enough to override the other cues in the S#N frame (i.e. the S, P and N) and subjects responded with more S#N responses for all levels of K that were inserted. These K's made a marginal difference, in that K#S responses would increase slightly with the insertion of a long K, but the majority response was still S#N.

The Effect of S

The S's clearly affected every frame. For all frames, inserting a long S elicited S#S or K#S responses, while inserting a short S elicited S#N responses. Intermediate length S's provided cross-over points between S#N, and K#S/S#S. In all frames, the S cue could override all three combinatory cues provided by the frames (i.e., K, P, and N). As in the first two experiments, the number of K#S responses was higher than S#S responses, indicating a subject preference for K#S. The number of responses in the S#N category, however,

was always high for appropriate S values.

However, in the K#S frame, the insertion of short S's led to a majority of S#N responses, but not to the same high levels that were found when these S's were inserted into the S#N and S#S frames. Presumably, the long K in the K#S frame interacts with the S cue in an inhibitory way. It is also important to note that the S can override all other cues (K, P, and N), but the K duration cannot.

The Effect of P

In the S#N frame, a long pause elicited more S#N responses. In the other two frames, a long pause increased the number of S#N responses slightly, but the majority of responses was determined by the frame. Therefore, no 'swamping' occurred and the pause did not override cues in the S#S and K#S frames. It may be that the longest pause was too short to elicit such response behaviour. This suggested that a wider range in pause duration would be necessary for the fully crossed experiment, in order to obtain cross-over points in the identification curves.

The Effect of N

The duration of the N seemed to strengthen the responses for the S#N frame. The other frames were affected only to a small degree by the length of the N. Thus, N by itself could not override the other three cues (K, S, and P) provided by the S#S and K#S frames.

In summary, the S duration could override all other cues and is therefore, considered a primary signalling agent of juncture. The K duration could override cues in the S#S frame, indicating that it is a signal for cluster K-S vs. non-clustered K#S. N's were not strong enough to override the S and K cues in S#S and K#S frames. Pause levels were too short, and the effects of P therefore are inconclusive. In general, the ranges of elements provided cross-over points (except for pause, whose range may have been too narrow). Thus, these ranges of elements could be used in the following fully crossed experiment.

**Rating Results**

Subjects were asked to cross off the number of the item if, after identification, they felt that the utterance was somehow odd for the selected category. Analysis of the results indicated that subjects marked off an utterance as being 'odd for that category' in two cases:

1. MISIDENTIFICATION: Subjects classified the utterances in a category of which the elements were typical of another category.
2. CONFLICTING CUES: The variable inserted was an extreme which conflicted with typical cues within the frame that happen to be extreme. (For example, inserting a long S in the K#S frame, which already has a long K.)

It is notable that subjects did not cross off typical frames which they 'correctly' identified.

When there were conflicting cues, and subjects crossed off these utterances as being odd, the following response patterns emerged. When long K's and long S's were combined, 35% of the subjects crossed off the utterance, and the S cue generally 'won'. However, there were differences among the subjects; some subjects responded with a K#S in this case, indicating that, for them, the K predominated. (Approximately 60% of those subjects crossing off these utterances responded with S#S, whereas the others responded with K#S.) When long S's and long P's combined, the P was a stonger cue for 70% of those subjects marking these as being odd combinations. About 35% of the subjects had crossed off these utterances. However, as noted by the original graphs (Fig 8), most subjects were influenced by the S. Thus, for subjects who thought these combinations were odd, the pause predominated, thereby yielding S#N responses. When long K's and long P's combined, again the pause affected listeners eventual choice (i.e., more S#N responses were found under this condition for the 10% of subjects who crossed off these utterances).

These results indicate that for combinations which sound 'odd' to subjects, later cues affect responses more than earlier cues, when these cues are in conflict. Admittedly, these results are hard to interpret since the relative salience among elements is unknown. More work is needed on the rating technique and it may prove to be a useful paradigm for eliciting information regarding the

finer detail of subjects' decisions.

## B. Experiment Four: The Fully Crossed Experiment

Experiment 3 showed that changes of single elements in a frame could alter perception, but not completely override all the frame cues in all conditions. Experiment 4 was designed in order to test the complete combinatory effects of the K, S, P and N segments. Since the P levels in Experiment 3 were in too narrow of a range for meaningful cross-over points, the range was increased for this experiment to 0-72 msecs. (as opposed to the 0-45 msecs. range in Experiment 3). All ranges from the other variables were similar to those in Experiment 3.

This combinatory experiment was designed in order to test what would happen if all of the elements changed. What combinations would lead to changes in juncture perception? How do listeners cope with 'conflicting cues'? Finally, is there an interdependence of elements or is the perceptual decision-making process based on the K, S, P, and N cues independent of one another? It was important, therefore, to cross levels of all four factors in order to address these questions.

### Subjects

There were 36 first-year psychology students who participated in the experiment for course credit. Two subjects were omitted because they marked their sheets in an improper manner. All subjects were native English speakers

and had no known hearing difficulties.

Lists of Stimuli

Owing to the large number of stimuli in such fully crossed experiments, only one randomization list was prepared. Four levels of S, and three levels of N, K, and P were fully crossed producing 108 combinations. The beginning portion SHLK and the end portion APPING were attached to each combination to complete the sentences. The stimuli and their duration values are listed in Appendix 4. These levels were picked because, according to Experiment 3, such duration levels would provide appropriate crossover points. In addition they were equally spaced in terms of linear duration. Each stimulus was repeated twice. After every five sets of repeated stimuli, a digitally prepared sine tone was inserted in order to allow subjects to keep track of their responses. The stimuli were recorded on audio tape.

Procedure

Subjects were asked to identify each sentence and mark out a 1, 2, or 3 on a computer-scorable answer sheet corresponding to S#S, K#S, and S#N, respectively. The 'Key' sheet of Experiment 1 was provided. A practice block of ten stimuli, taken from the middle of the stimulus list was played before the actual experiment. The subjects marked trial responses on a separate computer sheet.

The 108 stimulus list was broken into two parts, and the test was conducted in two sessions. In the first session the first 55 stimuli were played. After a break of ten

minutes, the second session was conducted in which the last 53 stimuli were presented. The test was conducted in this way in order to reduce subject fatigue effects. The stimuli were so divided in order to make the location of the sine tones coincide with the demarcations of the computer score sheet.

**Results**

The listeners' responses were totalled and the identification curves were plotted (Fig 9). The effects of each variable are discussed below.

The K-closure

The length of the K generally distinguishes clustered K's from the non-clustered case (i.e., tends to cue a K#S response). If the N is short and there is not a long pause, a large K yields more K#S responses. However, if the N is long or if a long pause is there, this does not take place. Thus, when K and N 'conflict' (i.e., long K signalling K#S and long N signalling S#N) neither element really dominates in the listener's judgement. Instead, the judgement seems to be more dependent on the S duration. When the K and P elements conflict (i.e., long K signalling K#S and long P signalling S#N) listeners' judgements again depend more on the S duration, especially if there is a shorter N.

The K, then, can be seen to have a greater effect when the other elements are short; a long K under these conditions will signal a non-clustered final K, thus

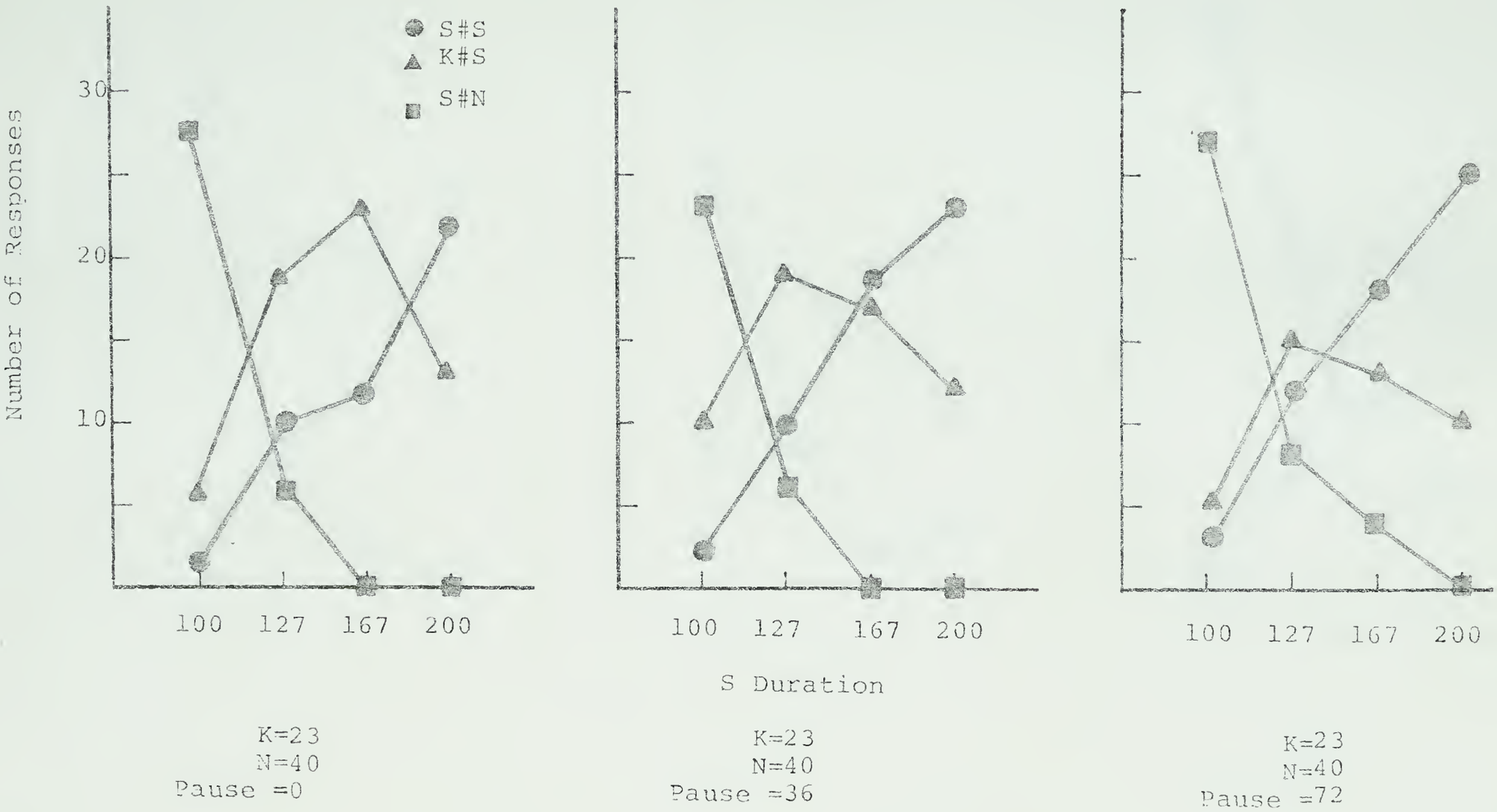



Fig 9. Crossed Experiment Curves

All durations are in msecs.

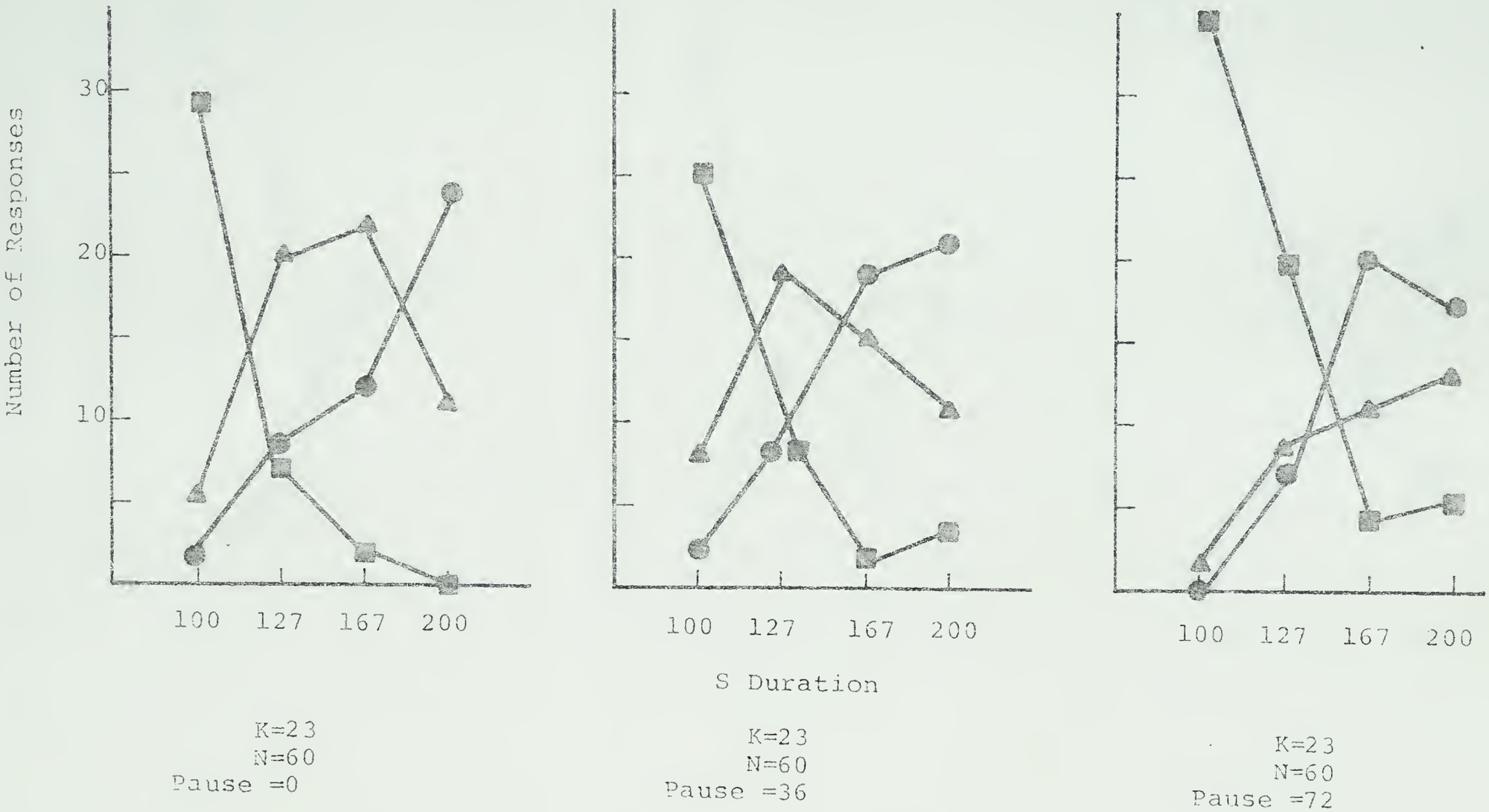



Fig 9. Crossed Experiment Curves
(continued)

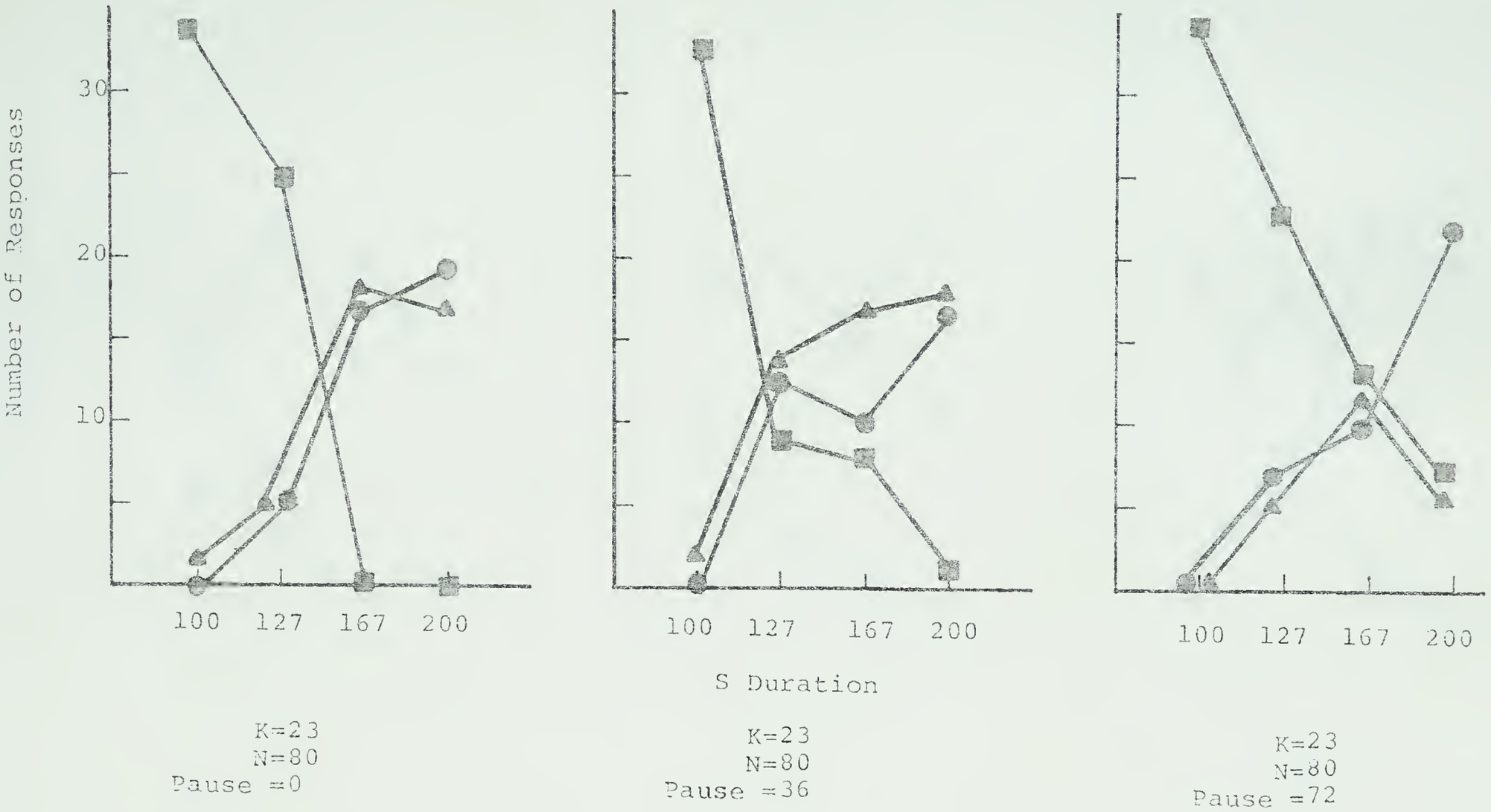
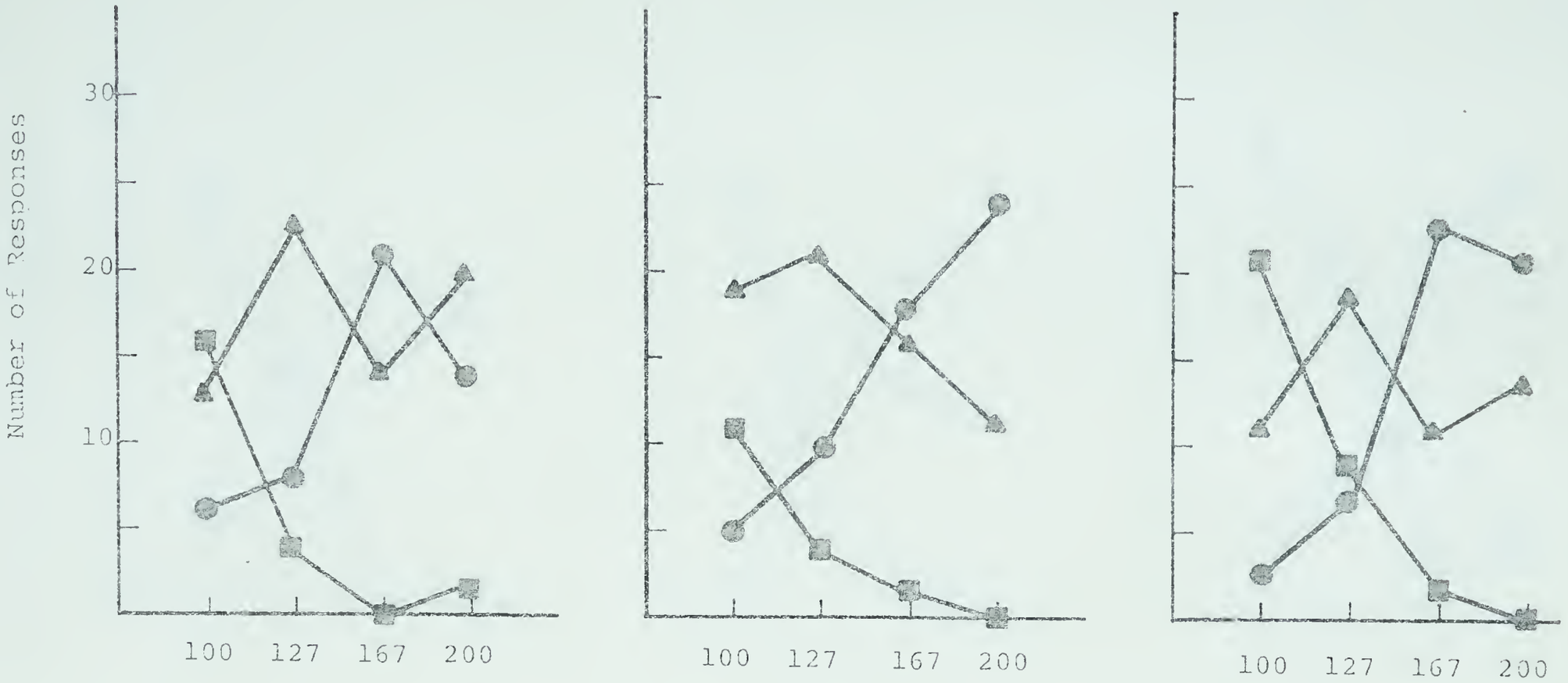



Fig 9. Crossed Experiment Curves
(continued)

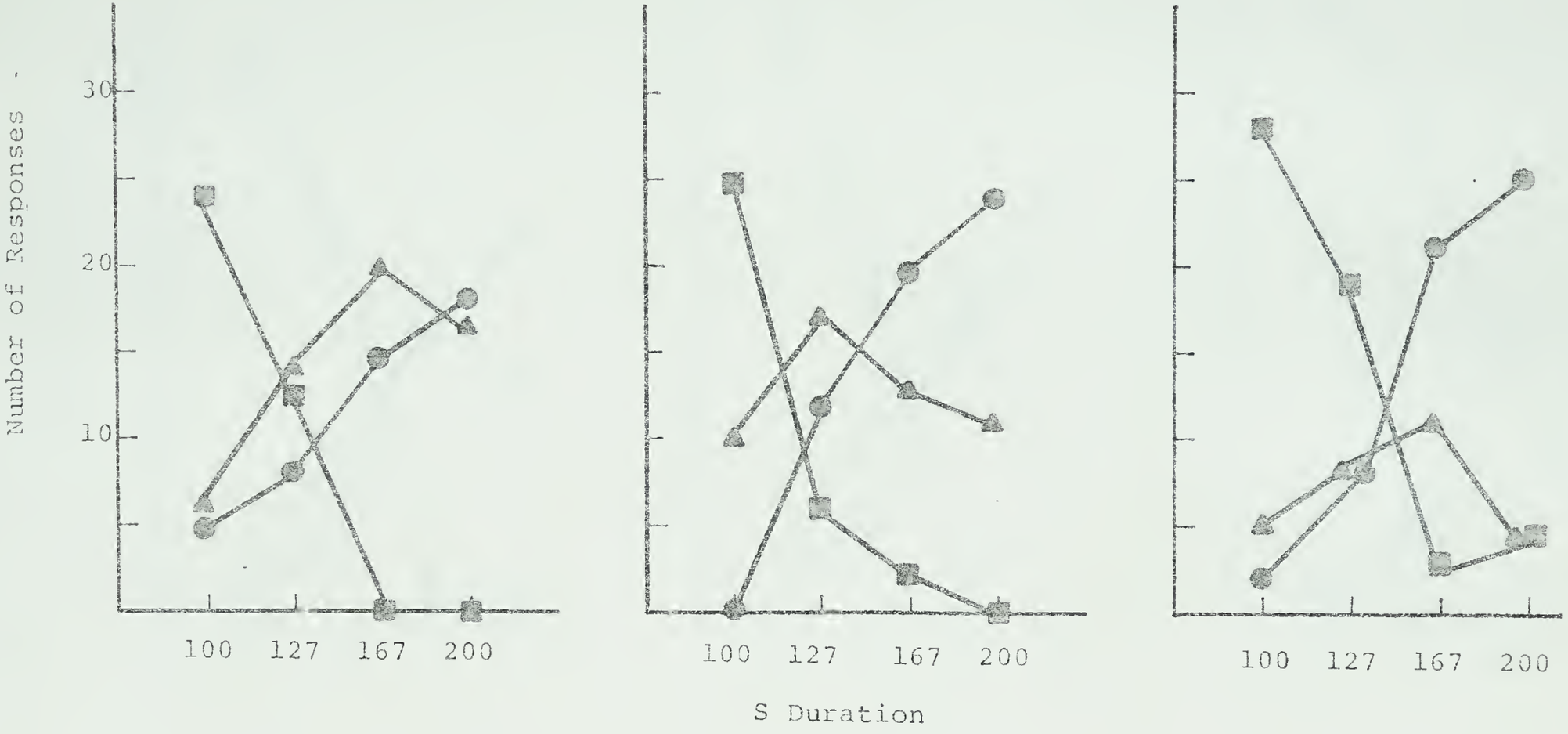



K=54
N=40
Pause = 0

K=54
N=40
Pause =36

K=54
N=40
Pause =72

Fig 9. Crossed Experiment Curves
(continued)

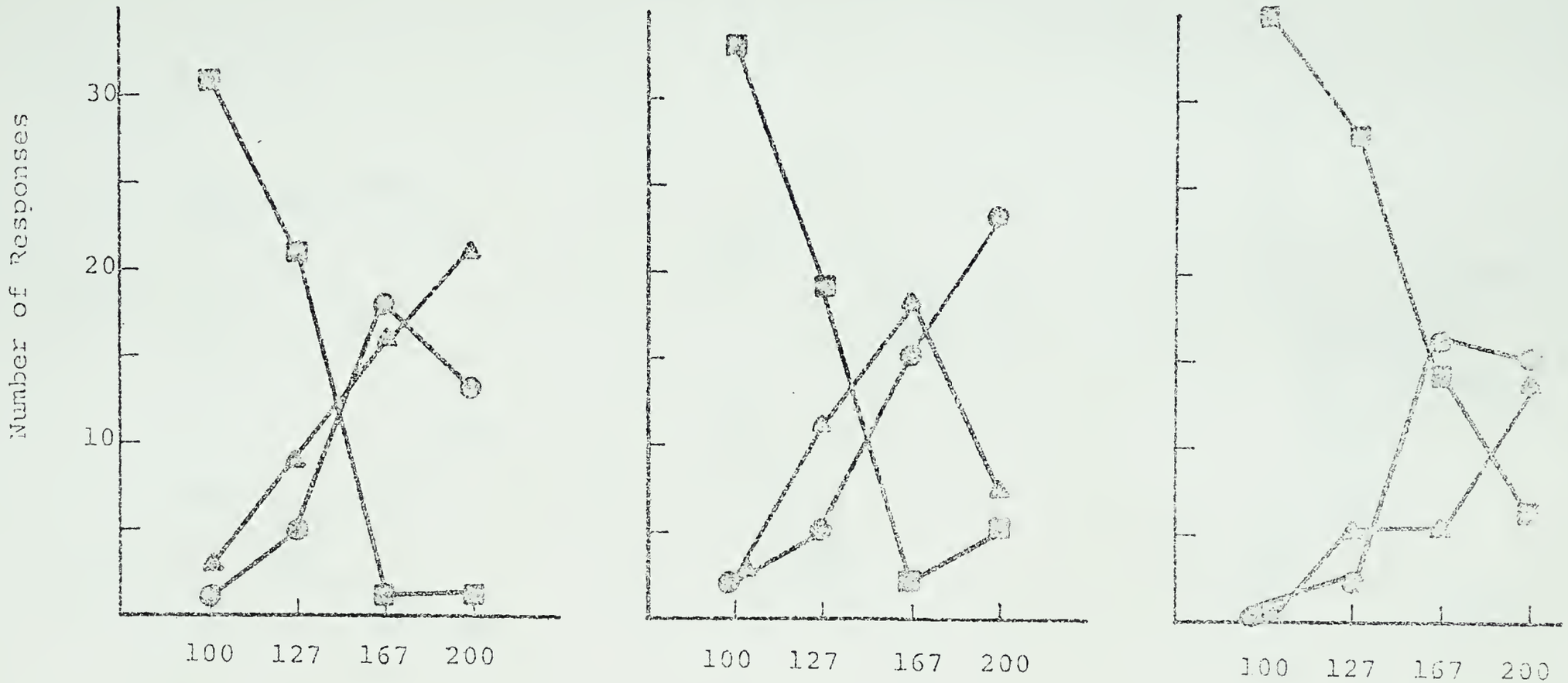



K=54
N=60
Pause =0

K=54
N=60
Pause =36

K=54
N=60
Pause =72

Fig 9. Crossed Experiment Curves
(continued)

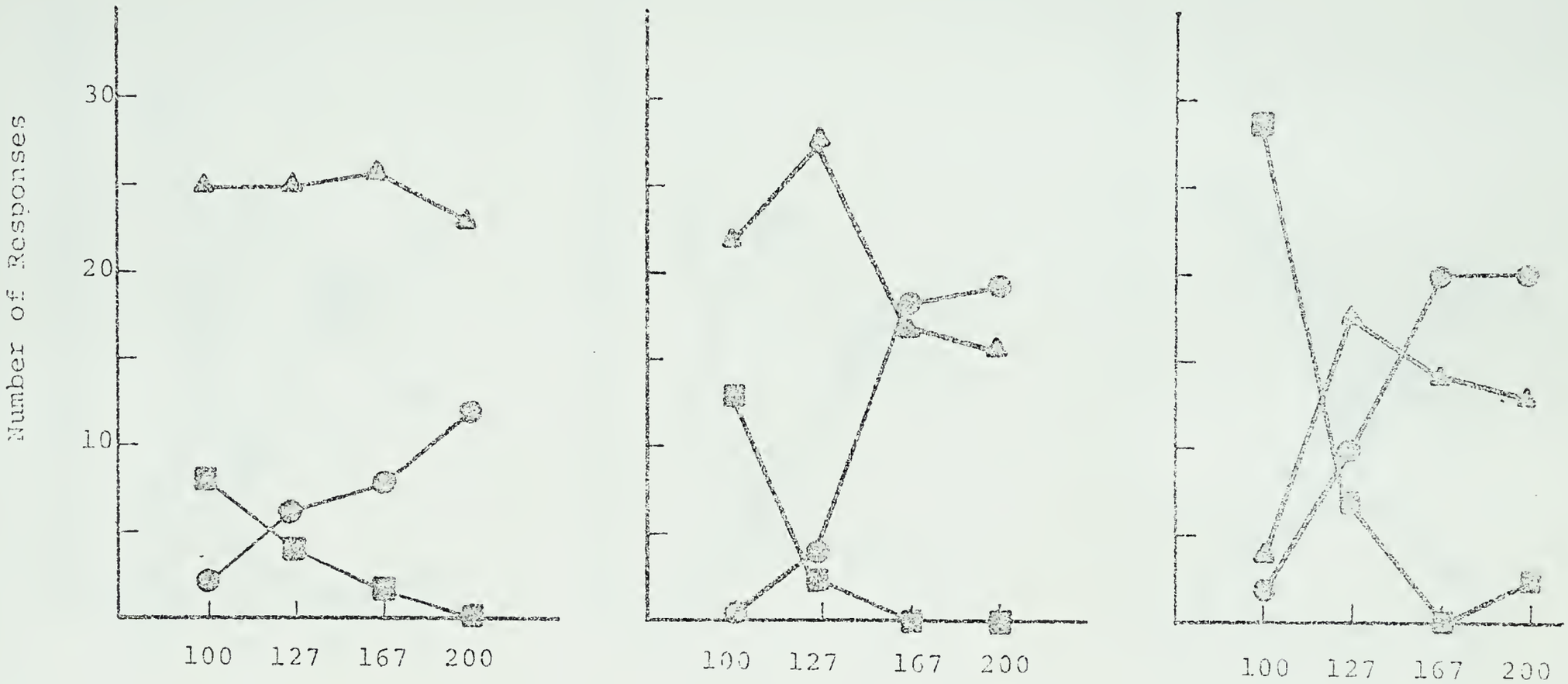



K=54
N=80
Pause =0

K=54
N=80
Pause =36

K=54
N=80
Pause =72

Fig 9. Crossed Experiment Curves
(continued)




K=85
N=40
Pause =0

K=85
N=40
Pause =36

K=85
N=40
Pause =72

Fig 9. Crossed Experiment Curves
(continued)

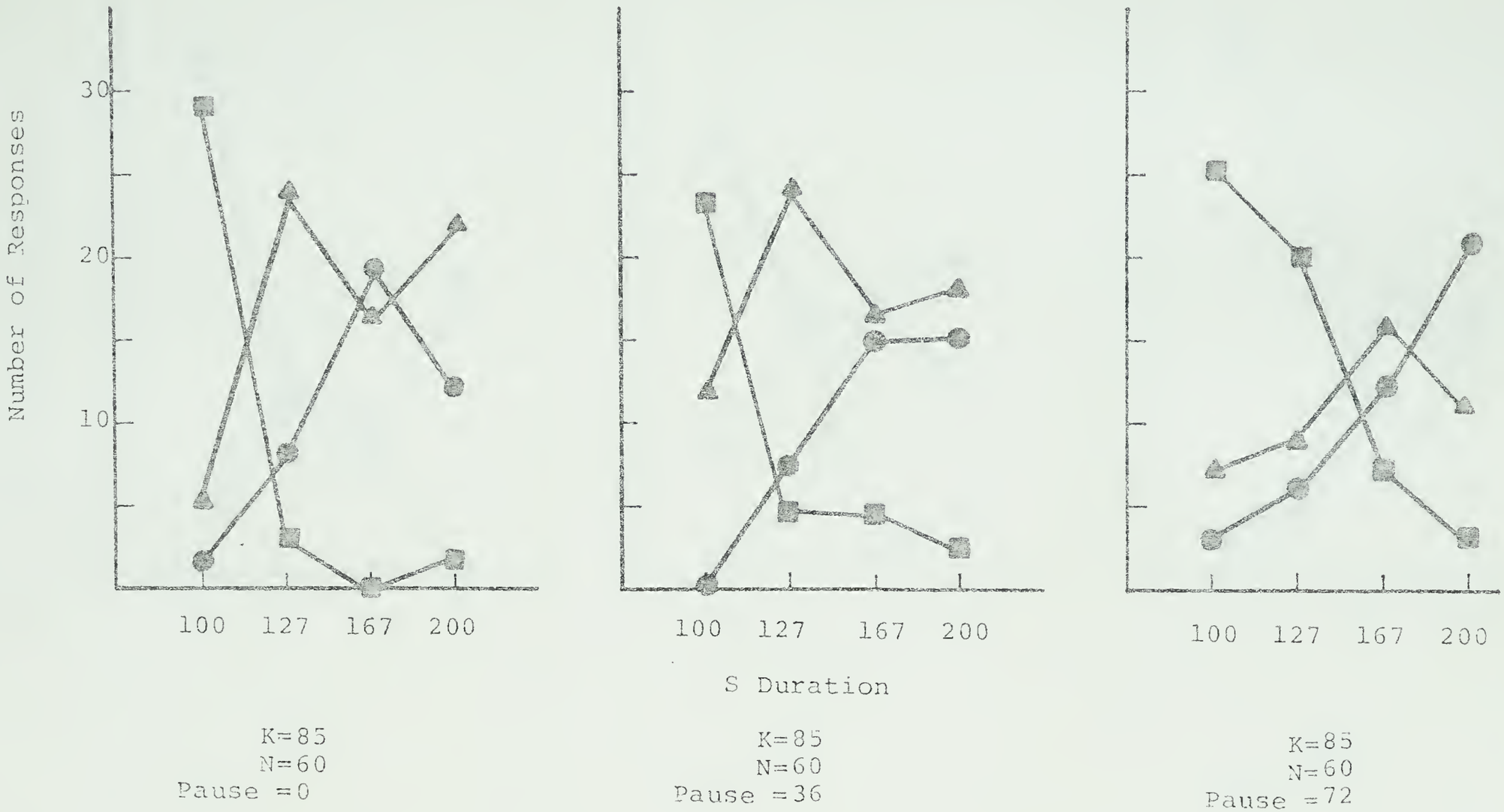



Fig 9. Crossed Experiment Curves
(continued)

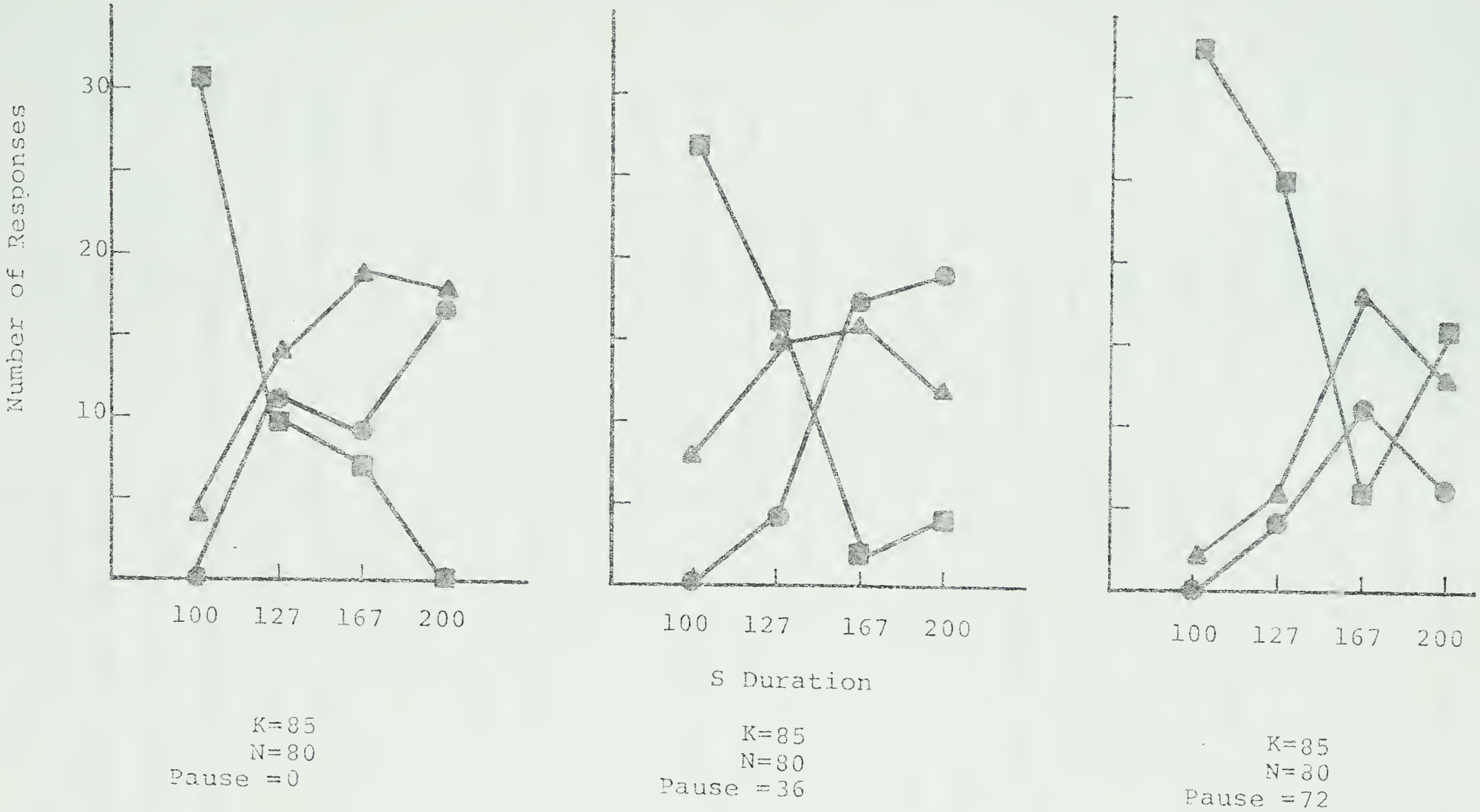



Fig 9. Crossed Experiment Curves
(continued)

yielding more K#S responses. However, when these other elements are larger, the K effect is less and longer K's do not elicit these high levels of K#S responses.

The S duration

The S duration is a powerful signalling cue. As expected from the swamping experiment, the duration of the S is a major cue for signalling juncture location; all graphs show that smaller S's lead to S#N judgements, larger S's lead to K#S or S#S responses, and the longest S generally cues the S#S condition. There were, however, a few exceptions which are outlined in the following paragraph.

As mentioned above, if the K was long and the other elements were short, the short S cue (signalling S#N) was overridden by the K cue, thus yielding more K#S responses. A second exception can be found when all elements are long. Here, the longest S signals more S#N responses than K#S or S#S. However, since the rest of the cues are 'conflicting' in this case, subjects seem to be relying on the latest signal. This was also evident in the swamping experiment; when utterances sounded 'odd for that category' (as seen by subjects' cross-off ratings) their judgements were based on the latest cues, namely the P and N, thereby yielding more S#N responses. Also, there are two cues here for S#N (namely, P and N) and only one for each of K#S and S#S; perhaps it is a cue majority and not a 'latest cue'

which makes this occur. (See perceptual models for segmentation in Chapter 6.)

Another interesting point should also be noted. While short S's invariably lead to S#N responses, longer S's can lead to either S#S or K#S responses. In some cases, particularly with a long N, the S#S and K#S curves are almost identical. Thus, the curves show that S#N behaves markedly different from the other two categories, but that the K#S and S#S curves are not different from one another. This is particularly evident when the long pause is used. This suggests that subjects could more easily judge a S#N category than distinguish between S#S and K#S (i.e., they could tell a final S had occurred, but were undecided when an initial S vs. a double S had ocurred). In this crossed experiment, a K#S preference was not found; rather S#S responses were highly evident, and in some cases, overshadowed K#S responses. Many subjects responded with S#S when the longest S was present; perhaps they developed an ad hoc strategy of listening directly for the long S in order to judge this category.

The Duration of the Pause

The effect of increasing the pause duration was that more S#N responses were obtained. In addition, when pause conflicted with other cues, it could not override the S cue except for the case, mentioned above, where every segment is long. The other effect of pause, as

mentioned above, was that the responses in a long pause condition showed identical K#S and S#S curves.

The N Duration

Increasing the duration of N strengthened S#N responses. In all cases where the long N was present, the K#S-S#N boundary was pushed to the right, indicating that more S#N responses were being given. In addition, very long N's would not allow long K's to override the S cue, as was the case when N's were short, indicating an inhibitory role.

## Statistical Analysis

The identification curves (Fig 9) were statistically analyzed in several ways. Owing to the general lack of experience in applying statistical models to such data, it was necessary to apply several statistical evaluations. The statistical models used included the ANOVA and several logistic models; these analyses were then compared to one another.

The ANOVA model is easy to apply since the perceptual experiments followed a fully crossed design. However, each category had to be analyzed separately; thus, while the data were treated as three separate problems, this was not in fact the case. An ipsitive relationship existed between the three categories - by knowing the responses in two categories, the third category responses were automatically known. In some cases, significant interactions were found that were suspect, owing to this ipsitivity; some

interactions were found only in one of the three categories, but owing to the ipsitive relationship that exists among these categories, it is expected that if an interaction was truly evident, it would be significant in at least two categories. In addition, these same interactions were not found to be significant when subjected to the logistic analyses.

The logistic analyses fit several specified models to the data. This procedure, in some respects, is a more justifiable model for such multi-categorical data, since an ipsitive relationship is built into the logistic model itself.

In the last section, the ANOVA and logistic models are compared. A model is chosen on the basis of converging results of the two analyses.

ANOVA Results

For each category, response proportions were fit into the arcsin space by the following formula

$$2*\arcsin\sqrt{(\text{proportion})+.25/\text{Total}+.5}$$

This was done for reasons discussed in Bock and Jones (1968, p. 72). ANOVA tests were done on each category separately so that significant main effects would be meaningful. However, as mentioned above, an ipsitive relationship exists among the categories, which is not taken into account by this ANOVA analysis. The ANOVA results are shown in Table 7.

For the S#S Category, the main effects of N, K, and

TABLE 7

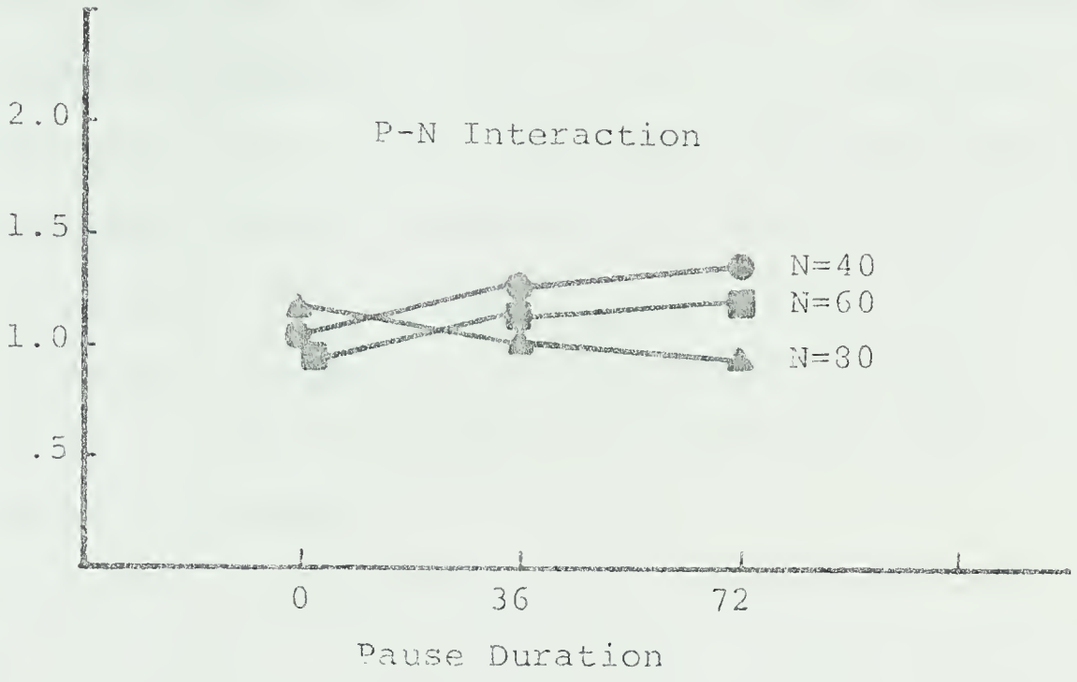
Anova Results for the Crossed Experiment

| Source | D.F. | F Ratio | | |
|---|---|---|---|---|
| | | Category S#N | Category K#S | Category S#S |
| N | 2 | 254.36** | 55.8** | 21.84** |
| K | 2 | 8.15** | 23.59** | 13.19** |
| P | 2 | 129.87** | 59.12** | 2.2 |
| S | 3 | 1094.23** | 77.54** | 373.27** |
| NK | 4 | 2.53* | 1.06* | .94 |
| NP | 4 | 4.42** | .44 | 4.66** |
| KP | 4 | 1.21 | 1.27 | .55 |
| NS | 6 | 12.76** | 13.28** | .35 |
| KS | 6 | 10.57** | 2.91* | 3.68** |
| PS | 6 | 4.45** | 4.73** | 1.87 |
| NKP | 8 | 1.48 | 2.01 | 1.31 |
| NKS | 12 | 1.64 | 1.15 | .94 |
| NPS | 12 | 4.03** | 1.51 | .95 |
| KPS | 12 | 2.28* | 1.5 | 2.18* |
| NKPS | 24 | 5.32** | 1.66* | 2.17** |
| G(NKPS) | 108 | | | |

S are significant. From the cell means it can be seen that shorter N's yield more S#S responses, while longer N's yield fewer S#S responses. The same trend is evident for the K's. This trend is reversed in the S-case; longer S's lead to more S#S responses, shorter S's to less. In addition, the PN, KS, and NKPS interactions are significant. The two-way interactions are shown in Fig 10. The PN interaction accounts for only a very low proportion of the total variance (1%) and, furthermore, is suspect in view of the ipsitivity problem, since it only showed significance in this one category. (Also, the logistic analysis (discussed below) found this interaction to be non-significant.) The KS interaction shows that long K's lead to fewer S#S responses when the S is short than when the S is long, in relation to the other K levels. Thus, when S is short and K is long, more K#S responses result (i.e., less S#S responses) but this same K does not have this effect when followed by a long S. The NKPS interaction follows the trends of the main effects; this interaction will not be further analyzed since it accounts for only a small proportion of the total variance ($<3\%$).

For the K#S category all main effects are significant. Cell means show that

1. the longer the N, the fewer the K#S responses, because long N's tend to elicit S#N responses
2. the longer the K, the more K#S responses



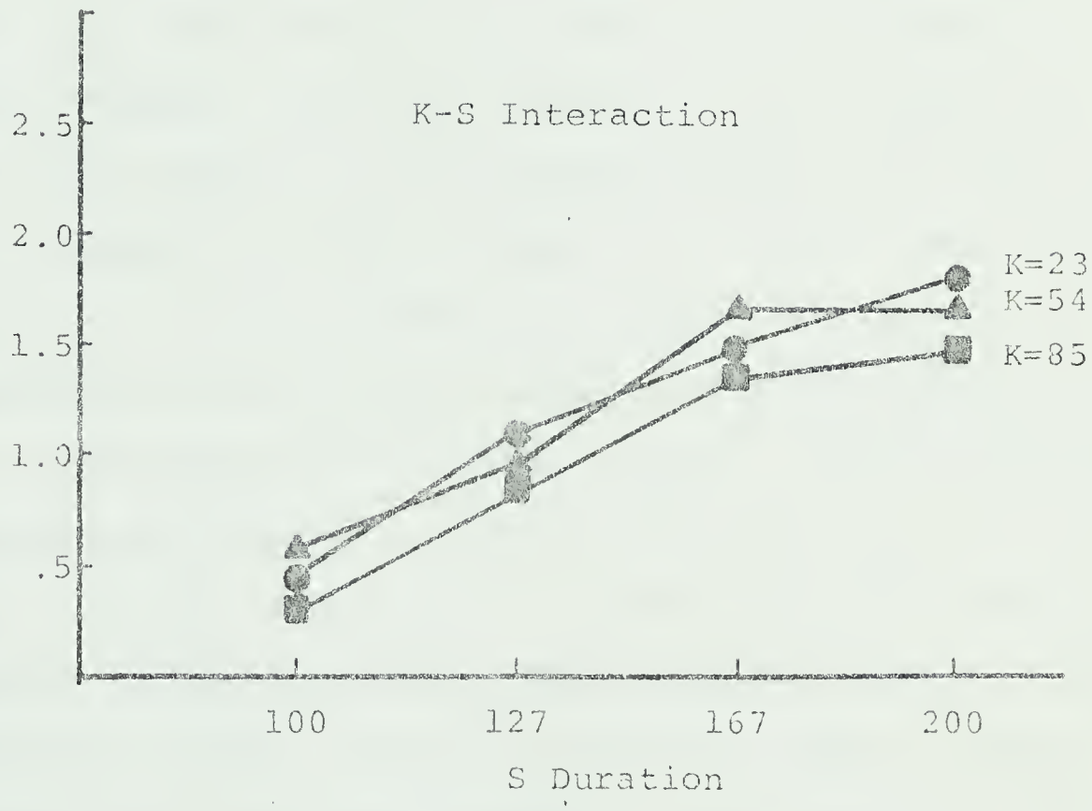




Fig 10. S#S Interaction Plots

3. the longer the pause, the fewer the K#S responses, because long pauses tend to elicit S#N responses
4. shortest S's (a cue for S#N) yield the least K#S responses, while intermediate S's yield the most K#S responses; the longest S leads to fewer K#S responses, because it is a cue for S#S. However, the level of K#S responses even for the longest S is still quite high.

The SN and PS interactions were significant and are shown in Fig 11. In the SN case, longer N's tend to elicit fewer K#S responses when preceeded by a short S (S#N cue) than when these same N's are preceeded by a long S. However, short N's and short S's produced nearly the same number of K#S responses as short N's and long S's; presumably the S#N category is taking away some responses in the former case, while S#S responses are predominant in the latter case. For the PS interaction, as S increases with no pause present, the K#S response becomes more frequent. However, this is not the case when a pause duration is present; when the pause is present, subjects do not produce a greater number of K#S responses as the S duration increases. When no pause is present, the S cue acts alone, but when pause is there, a conflict arises between S and pause.

In the analysis of the S#N category, all main effects were significant. Judging from cell means,

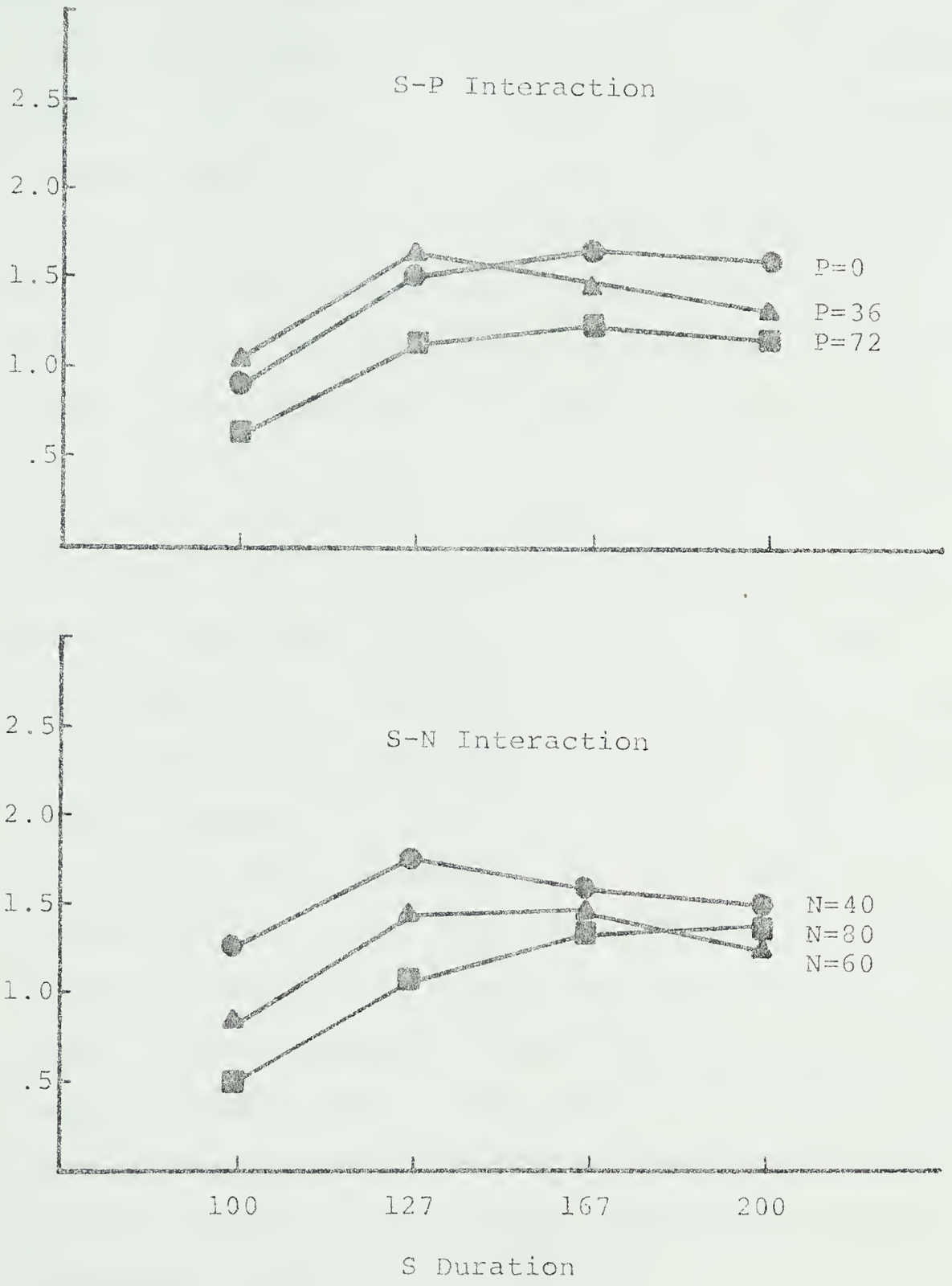
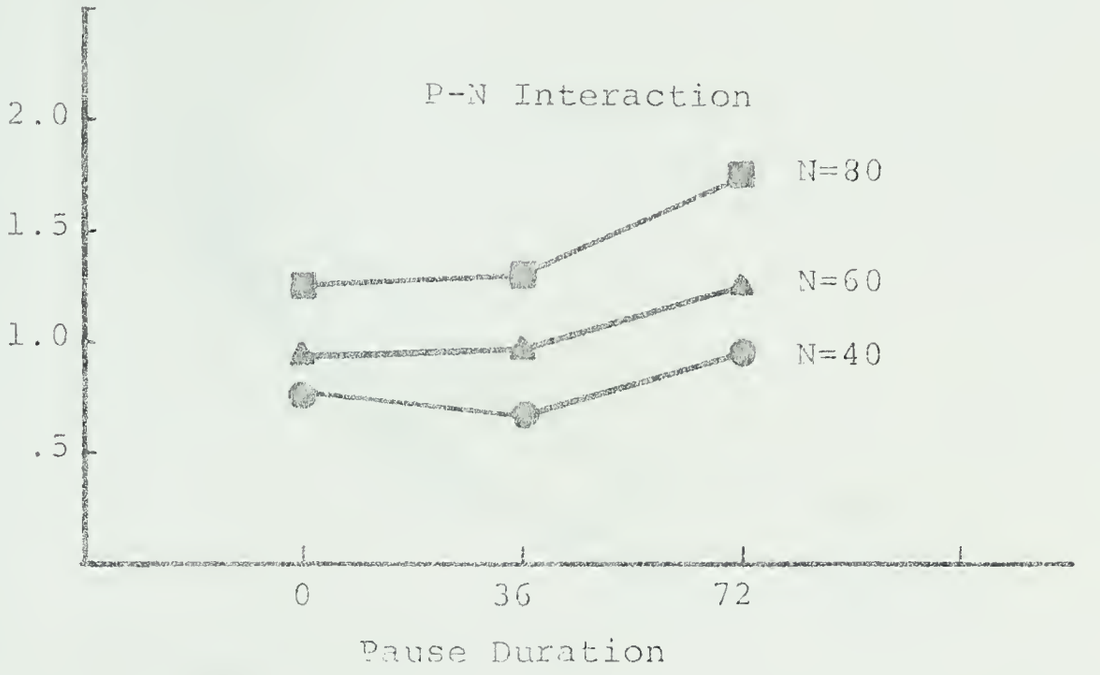



Fig 11. K#S Interaction Plots

longer N's and pauses yielded more S#N responses, while shorter K's and S's led to more S#N responses. Several significant interactions were evident for this category, namely, PN, SN, KS, SP, NPS, and NKPS. The two-way interactions are plotted in Fig 12. The three and four-way interactions follow the main effects trends and will not be further analyzed, since they account for small proportions of the total variance (.07% and 2%, respectively). For both the PN and SN cases, it can be seen that the trends of the main effects are evident, but only a difference in degree exists between the different levels of each of the two factors. It could be the case that these interactions are found when this measurement scale (arcsine transformed counts) is used. The KS interaction also shows the basic main effects trends. However, it seems that short K's and short S's produce more S#N responses than when these K's are combined with longer S's, in relation to the other K levels. In fact, there seems to be no difference in the number of S#N responses between the three levels of K at longer S values. The SP interaction in this category seems to be the inverse of the SP interaction in the K#S category; there is a difference between no pause and some pause in relation to S levels.

Problems With ANOVA

Unfortunately, statistical results of categorical data using the ANOVA tests have several difficulties.



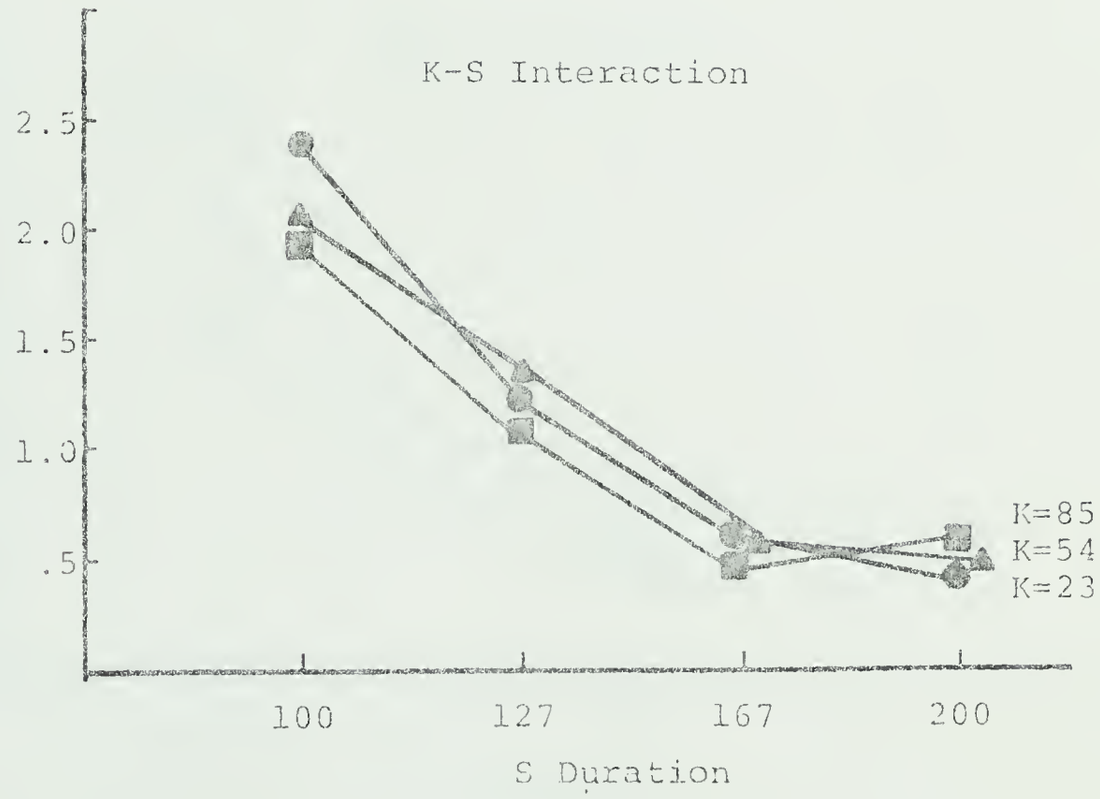
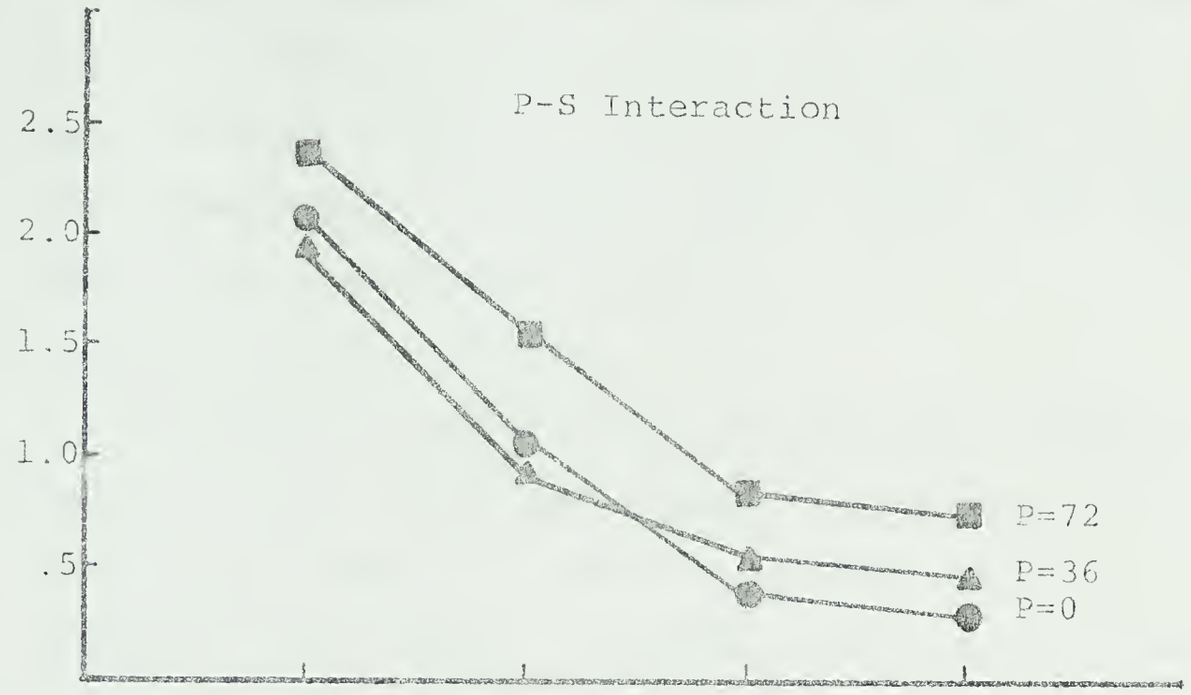




Fig 12. S#N Interaction Plots



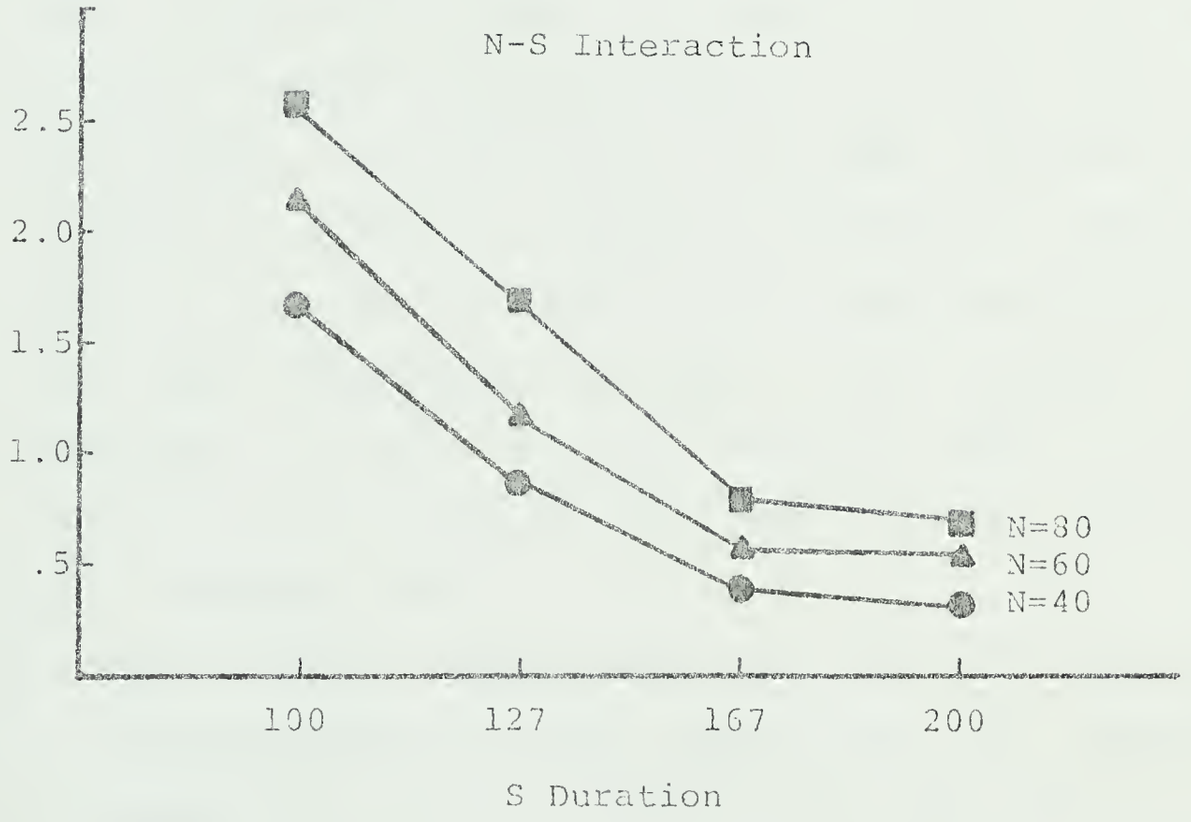




Fig 12. S#N Interaction Plots
(continued)

First, the ANOVA assumes homogeneity of variance, which in categorical data is not necessarily true. Second, there is the ipsitivity problem (discussed above). The basic problem, then, is that we are dealing with multi-dimentional frequency tables, where the ANOVA model is an inappropriate statistic.

The Logistic Model Analysis

The logistic analysis is a fairly new method of dealing with multi-dimentional contingency tables. Statistical tests in this model are based on the assumption that the set of responses to a given stimulus are drawn from a product-nomial distribution (Fienberg, 1978, Nearey and Hogan, to appear). Thus, the three categories can be discussed together, and their ipsitive relationship is built into the model. This data was analyzed in terms of several hierarchical logistic models which are outlined below, and were done by a log-linear programming package. The results are tabulated in Table 8. The statistic used in this analysis is the so-called likelihood ratio chi squared, $G^2$ (Fienberg, 1978).

1. Main Effects Model (CS, CK, CP, CN): Only the interactions of the main effects with the juncture categories, with no interactions among the K, S, P and N elements are in this model. This model appears to provide a reasonably good fit to the data and is a good first approximation to the identification

TABLE 8

Logistic Results

| Terms Included | $G^2$ | D.F. | $\Delta G^2$ | $\Delta G^2$ | Significance |
|---|---|---|---|---|---|
| GKSPN | 2218.02 | 432 | | | |
| CK,CS,CP,CN | 376.63 | 412 | 1841.39 | 20 | $>.005$ |
| CKS,CSP,CPN | 312.90 | 380 | 63.73 | 32 | $>.005$ |
| CKP,CKN,CSN | 290.18 | 352 | 22.72 | 28 | n.s |
| CKSPN | 131.96 | 216 | 158.22 | 36 | n.s. |

curves of Fig 9. However, the global goodness of fit test cannot be taken at face value when many of the expected values are small (Haberman, 1978). [4]

2. Adjacent Two-way Interactions Model (CS, CK, CP, CN, CKS, CSP, CPN): This model has the main effects plus the interaction terms which are immediately adjacent temporally (ie. KS, SP, and PN). This model was significantly better than the Main Effects Model; the goodness of fit to the curves of Fig 9 significantly improved.
3. All Two-way Interactions Model (CS, CK, CP, CN, CKS, CSP, CPN, CKP, CKN, CSN): This model was based on the main effects plus all two-way interactions. It was not significantly better than the Adjacent Two-way Interactions Model. Therefore, adding those two-way interactions which are not temporally adjacent does not provide a better goodness of fit to the data.
4. Higher-order Interaction Models (CKSPN): Adding three or four-way interactions into the model did not significantly improve the model's goodness of fit, as compared to the fit obtained by the Adjacent Two-way Interaction Model.

---

[4]Note, however, that Haberman (1978) indicates that tests based on *differences* in $G^2$ are generally close to their descriptive significance levels when comparing models, even in this case.

In order to test which of the Adjacent interactions were important, a 'backwards stepwise' analyses was done, using procedures described in Fienberg (1978). In this procedure, the model containing all Adjacent interactions is compared with models whereby one of these interactions is dropped at a time. Results are tabulated in Table 9. It was found that dropping the PN interaction did not produce a significant difference in the goodness of fit, but dropping either the KS or PS lowered the goodness of fit significantly. As a second step, a Modified Adjacent Two-way Interaction Model was tested which had the main effects plus only the KS and PS interactions. Dropping either the KS or PS from this modified model again significantly reduced the goodness of fit.

Comparison of ANOVA and Logistic Analysis

Although it is difficult to compare the ANOVA and logistic analysis, it is possible to at least examine the differences between these two methods of analyses. In the following discussion, it is noteworthy to keep in mind that the ANOVA was conducted on each category separately while the logistic analysis combined all three categories. The other major difference between both methods is the nature of the analysis. The ANOVA is a model with additive effects, whereas the logistic is based on effects which are basically multiplicative in nature. Therefore, an exact comparison between the two

TABLE 9

Backwards Stepwise Analysis

STEP 1 BASE MODEL = CKS,CSP,CPN

| | | | Difference from Model | | |
|---|---|---|---|---|---|
| Effect Removed | $G^2$ | D.F. | $\Delta G^2$ | $\Delta$D.F. | Significance |
| CPN | 318.27 | 388 | 5.37 | 8 | .70 |
| CSP | 338.59 | 392 | 25.59 | 12 | .02 |
| CKS | 343.35 | 392 | 30.45 | 12 | .01 |

STEP 2 BASE MODEL = CKS,CSP

| | | | Difference from Model | | |
|---|---|---|---|---|---|
| Effect Removed | $G^2$ | D.F. | $\Delta G^2$ | $\Delta$D.F. | Significance |
| CKS | 345.35 | 400 | 27.08 | 12 | .001 |
| CSP | 349.31 | 400 | 31.04 | 12 | .001 |
| CN | 583.33 | 392 | 265.04 | 4 | |

methods is difficult. The two analyses can be summarized as follows, using only the main effects as an example:

The ANOVA Model:

$$Y(spknr)=m+a(k)+b(s)+c(p)+d(n)+e(kspnr)$$

where the error term $e(kspnr)=N(0,\sigma)$.

In this model, three separate analyses are done for each category.

The Logistic Model:

$$Y(CKSPNR)=\frac{\mu\ [A(CK)\ B(CS)\ C(CP)\ D(CN)\ F(C)]+E(CKSPNR)}{[A(jK)\ B(jS)\ C(jp)\ D(jN)\ F(j)]}$$

where $A(CS)=e(\exp\ \alpha\ CS)$

and E is a 'product multinomial' error distribution (Feinberg, 1978). Note that Y is constrained to range from 0 to $\mu$.

In this model, the three categories are built into the model simultaneously. The ANOVA model is largely inapropriate for these data, since it cannot handle the ipsitive relationship that exists between the categories. It also produces more significant interactions than the best fit logistic model. Although the logistic assumptions are not violated, there are still problems with this model, particularly with respect to its handling of low expected values. However, it seems to be the more appropriate model and yields less significant interactions; this is particularly appealing since those interactions which are not found to be significant in in the logistic model, but were so

in the ANOVA, were those very ones which were suspected of resulting as a consequence of the ipsitivity problem (cf. Snedekor and Cochrane, p.494, on differences between transformed probabilities and logits). [5]

In conclusion, all analyses point to the main effects and the the KS and SP interactions as being the only significant factors. It is noteworthy that the ANOVA model may produce additional significant interactions in categorical analyses. Future research on appropriate statistical models is necessary and it is suggested that for future experiments with a categorical structure, a more thorough analysis using logistic models is appropriate. Furthermore, strict reliance on the ANOVA model seems ill advised for an analysis of categorical data.

---

[5] T. Nearey(personal communication) conducted several pilot statistical experiments regarding the differences between the ANOVA and logistic analyses. Initial results indicate that these interactions are indeed a result of the ipsitivity problem.

## VI. Perceptual Models and Discussion

It was the purpose of the experiments to investigate bottom-up processing; that is, the amount of information that could be derived from the acoustic signal without syntactic, semantic, or pragmatic aid. It is not intended to suggest that such higher level processing does not influence speech perception; indeed, several previous investigations have led to the conclusion that such higher level processes are involved (Warren, 1970 and Morton and Long, 1971). Rather, this investigation deals with processes which are used in the absence of higher order information. It seems reasonable to assume that where these bottom-up processes are operative, higher order variables may either strengthen or weaken them, but such 'lower-level' information does play a role in perception of speech. Thus, listeners must gain some information from the acoustic signal as a 'first step' in understanding speech.

Many models of speech perception rely heavily on a top-down component. However, some models have recently been developed which depend more heavily on low level acoustic information. The first section of this chapter reviews several such models and outlines possible problems associated with their assumptions. The second section describes two alternative bottom-up processing models designed to account for word segmentation in speech. The concluding section suggests several experiments and outlines guidelines for their rationale.

## A. Some Bottom-up Models of Speech Perception

Bottom-up models of speech perception typically utilize aspects of the incoming signal to decode information and then send this information to higher level processes. The major dilemmas in such models are

1. what are the minimal units of the speech signal that may be attended to in speech perception
2. what are the low level storage limitations of memory and
3. at what stage is a decision made?

Wickelgren (1969) proposed that context-sensitive allophones were the minimal units in speech perception. The surrounding context was encorporated into these 'allophones'. Typically, a word such as 'stop' would be represented, in Wickelgren's terms, as $_{\#}s_{t}\ \ _{s}t_{o}\ \ _{t}o_{p}\ \ _{o}p_{\#}$ . This procedure tried to take care of a segmentation problem long inherent in speech research, as well as to give tentative ordering constraints on phonemes. The basic problem with Wickelgren's units is that they suggest that listeners will have *psycho-linguistically relevant* differences in perception between, for example, the /a/ in '$_{\#}a_{p}$' vs. the /a/ in '$_{p}a_{p}$' . While there is a distinct difference in the whole waveform between /ap/ and /pap/ sequences, there is no real evidence to suggest that a linguistic distinction is based on these differences; listeners will categorize both /a/'s as being /a/'s. While it is a matter for psychoacoustics to determine whether

listeners hear a difference between these two sounds, it is not clear that such a difference has linguistic relevance. Lehiste (1972, p. 188) commented on this distinction between auditory and linguistic processing.

> One of the problems in trying to establish what constitutes the minimal unit of speech is drawing a boundary between the perception of signals in a psycho-acoustic experiment (auditory processing) and the perception of signals in a speech mode (phonetic processing). It is well known that an identical physical stimulus may be perceived in two different ways, depending on the psychological setting...The question is now whether listeners are capable of distinguishing subphonemic phonetic detail while listening in a speech mode.

Lehiste (1972) also criticized Wickelgren's model on the basis that context in contextual allophones should be equally recoverable, while her experiments showed that this was not the case.

Several bottom-up models have been proposed for speech perception that attempt to define basic 'units of perception' which listeners attend to in the speech signal. These models were designed largely to account for a phenomenon known as 'categorical perception', in which listeners are more able to distinguish acoustic differences *between* labelled categories than *within* categories. For example, a particular 'b' will be more distinguishable from a 'd' than another 'b' even though the acoustic differences between all stimuli are commensurate.

A memorial model designed to account for this phenomena was suggested by Crowder and Morton (1969) and further developed by Fujisaki and and Kawashima (1969) and Pisoni

(1971). This model suggests that acoustic and phonetic information is differentially stored in memory; detailed acoustic information is stored for short periods of time until a phonetic categorization is done. At this point all irrelevant acoustic information is lost from 'precategorical acoustic' memory and phonetic memory retains only the linguistic classifications of these sounds. The phonetic store retains this classification for a longer time, but all finer acoustic detail has been lost. This model has not been further developed in order to handle boundary phenomena, but it is a general model that can be used in conjunction with such phenomena (see models for juncture perception, below).

Other bottom-up 'categorization' models encorporate feature detectors (Stevens, 1975) which, in an all-or-none response fashion, respond to parameters of the speech signal. Such detectors have been proposed for VOT, for example, on the basis of adaptation experiments in infants (Eimas, et al., 1971). Oden and Massaro (1978) attempted to show how continuous-valued featural information might be integrated in speech perception. Although there may be problems associated with feature detectors, these models nevertheless point to interesting ways in which the perceptual mechanisms might operate in human speech reception systems.

Recently proposed models proposed by Nearey and Hogan (to appear) attempt to relate the categorical nature of perception to choice models based on distributions of

natural speech. They offer two basic models, the 'Threshold Thurstonian Model' and the 'Normal A Posteriori Probability Model' (NAPP). The Thurstonian model bases perceptual choice on boundaries between relevant categories, while the NAPP model bases perceptual choice on relative strength of group membership of the signal to the perceived category. They give examples of VOT classification experiments where subjects identify a signal as a 't' or 'd' based on a VOT attribute. For the NAPP model, they propose that such perceptual processes may involve a 'segment likelihood estimator' (SLE). The SLE's are "tuned so that their outputs are proportional to the probability density of their corresponding segments in the relevant population of signals in the language" (Nearey and Hogan, to appear). Therefore, listeners 'know' the distribution of speech sounds in their language. The SLE's are not an all or none decision in themselves, but merely encorporate the first stage of decisions, namely, the likelihood estimation stage. The second stage involves a choice mechanism such as that described by a Luce choice-theory model and which is encorporated into their NAPP framework. These models lead us to some viable psychological interpretations and are testable in experimental situations. What is required is to examine if in fact listeners have knowledge of the distribution of natural speech sounds in their language. This will be discussed later with regards to implications for future research.

## Models from Research in Computer Speech Recognition

Researchers involved in computer recognition of speech also developed some bottom-up aspects in their models of speech recognition. That is, they developed systems which gleaned information from the acoustic signal for higher level decisions. In most cases, phonemic recognition followed (see Wolf, 1976 for review). Klatt (1979), however, passed over this 'phonemic-decision' stage and fed acoustic parameters directly into a 'word-decision' process based on a best-match scheme (LAFS, or Lexical Access From Spectra). Klatt's model also differs in the use of basic parameters obtained from the acoustic signal. For example, HARPY (Reddy, 1976) segments the continuous waveform on the basis of zero-crossings and peaks, and a mechanism identifies features such as voicing, silence, frication, peak and dip detection, etc. These elements are then matched with 'allophonic templates' to provide phonetic labelling. In Klatt's model, however, spectral sequences are compared with internally represented 'diphones', or transition templates representing the connecting elements between two phones (called SCRIBER). These are similar to Wickelgren's allophones reduced to account for one context instead of two (ie. Wickelgren's $_{\#}a_{p}$ $_{a}p_{i}$ $_{p}i_{\#}$ = Klatt's $_{\#}a$ $_{a}p$ $_{p}i$ ...) This allows for template sharing in those sequences involving partially similar environments. As with Wickelgren's model, Klatt's model is at the disadvantage in proving a psycholinguistically relevant basis for his

diphones; in his other approach, LAFS, he must show psychological relevance for the word as a basic unit. Lehiste (1972) has questioned the relevance of this unit.

## B. Models for Juncture Perception

Most of the bottom-up models of speech perception have dealt with the special case of categorical perception for contrastive phonemic categories in a fixed context and have not dealt with boundary problem phenomena. It is of interest to devise models of juncture perception; while the computer recognition models have focused more on this problem, the psycholinguistic validity of those models is in question. A bottom-up model which has two versions is described below: the pairwise-evaluation model (PE) and the segment-allophone model (SA). The segment-allophone model is considered as a special case, or subset of the pairwise-evaluation model. These models differ in the choice of a basic unit of speech perception and in the procedure for word boundary segmentation.

The SA model uses the allophones of segments as minimal units; these segments are stored in memory and then a global boundary decision is made. The PE model uses adjacent pairs of segments as a special unit of juncture perception; these units are kept in memory, and the boundary decision is based on these segment pairs. The models lead to different mathematical treatments, which will be outlined below. Both the models considered here have two stages of decision. In

the first stage, signal properties are converted into phone-like strings. The *exact* nature of this conversion is not detailed. In the second stage, the output of the first stage is examined and 'corrected', if necessary, to yield a transcription consistent with English phonotactics.

The perception experiments indicated that listeners can segment an utterance in different ways depending upon the nature of the acoustic signal; they did not indicate which 'basic unit' was attended to nor did the experimental results lead to specific decision models. (In fact, in the statistical analysis of both the measurement and perception data, the three categories were treated as three separate unanalyzed categories. If such an approach were taken for a perceptual model, the basic unit of perception would be the phrase or sentence [eg. LIKES#NAPPING]. Clearly, this is unnacceptable. It is more appropriate to suppose that there is an early transformation from the segment to a linguistic representation.) These juncture perception models, then, are tentative proposals which require future experimental validation or refutation. They are presented in the hope of spurring experimental interest in word boundary phenomena. They are merely attempts at a 'first sketch' and by no means are intended to cover all aspects and problems in segmentation. Rather, they are designed in order to provide a conceptual framework for modelling decision procedures in juncture perception and will be generalized with the utmost caution. Finally, these models are proposed in order to

generate new questions regarding juncture phenomena and are by no means intended as answers to this very complex problem.

In providing a framework for juncture perception, two signal possibilities must be taken into consideration:

1. there are no conflicting cues in the signal and
2. there is at least one cue which conflicts with another cue.

The proposed models will both be discussed in these conditions.

**No Conflict in Cues**

When no conflict in the signal is present, pure bottom-up processing is possible. It is in this case that the perceiver can obtain information directly from the signal, without phonotactic or other higher-order input (except for the 'pairwise' phonotactics built into the first decision stage of the PE model). Under optimal communication conditions, the speaker would enunciate clearly, providing no conflict in cues; the listener would be able to decode this information rapidly and be more certain that he/she understood the utterance correctly. Thus, reaction time experiments should show that listeners are faster for these cases than when cues conflict in the signal. If listeners were asked to rate utterances as to their typicalness of that category, they should rate utterances with no conflicting cues higher than utterances in which cues conflict. Both versions of the proposed model handles this

case in which no cues conflict. This is discussed in the following sections.

The Pairwise-evaluation Model

This model assigns a special role to adjacent pairs of elements. Since interactions in the perceptual experiment (Chapter 5) were largely associated with adjacent segments (KS and PS), it may be worthwhile to consider a process that effectively deals with two elements at a time. In this model, subjects decide whether any two adjacent elements should be considered as a 'segment-pair', or as separate elements. For example, subjects could decide, upon hearing a short K and short S, that these two elements must both belong to the same word, and are therefore a cluster or fused segment-pair. The direct relationship between the attributes of the segments and the judgement of segment-pairs is given by the phonotactic constraints of the language (for example, measurements in Chapter 3 showed that the cluster at the end of the word LIKES had short K's and short S's). Presumably, classes of segment pairs could be defined (eg., [voiceless fricative+voiceless stop]) with generalized characteristic distributions. [6] Decisions as to these pairwise segments are stored in short term memory. A

---

[6]Exeptions to these generalizations must be included in the model; for example, [voiceless fricative+sonorant] is permissable except for these cases: *FW, *SR, *OW, *OL, (*ŠW), (*ŠL). In some cases, it is easier to state the exact element in the generalized case, for example in the case of [S+stop].

'second-stage' decision is made on the basis of all confirmatory information. In this case, when there are no conflicting cues in the utterance, this decision is easy; subjects choose the juncture location that is given by non-conflicting pairwise cues. An illustration is shown below, where the examples are taken from the perception experiments in Chapter 4.

Example 1: The PE Model with No Conflict in Cues

Suppose the pairwise decisions are as follows:

K--S, which suggests that the K and S are clustered, and

-S N:, which suggests that the S and N are not clustered.

This yields a tentative transcription of

[K-S][-S#N]

.

In this case, there is no conflict, and the word segmentation response which will be given is the LIKES#NAPPING response, since it is only this response in which the K and S are clustered and the S and N are not.

Subjects can not be expected to respond with 100% clarity as to whether or not they have heard a cluster (or segment-pair) vs. a non-cluster (or separated elements). To account for this, we might postulate probabilities associated with each pairwise decision. For example, suppose the following probabilities were obtained:

$$\left\{\begin{matrix}[K\text{--}S](.6) \\ [K\#S](.4)\end{matrix}\right\}\left\{\begin{matrix}[S\#N:](.9) \\ [\#S:N\text{-}](.1)\end{matrix}\right\}$$

This marking convention indicates that all possible segment pairs are evaluated (in this case there are four possibilities). Suppose also that subjects decided that a pause coincided with the S#N decision, with a probability of (.9). This would effectively raise the probability of this decision (eg., by 90%) and effectively lower the alternative pairwise desision (S-N), presumably by the same amount. Notice that the pause merely strengthens the probability, and is therefore considered as a separate process than the pairwise decision. [7] The highest average probability from all pairwise decisions yield a tentative phonetic transcription; if this is consistent with the phonotactic constraints of the language, as it is in the case of no conflict among cues, that is the response subjects give. Therefore, in this case, a LIKES#NAPPING response would be given the most often, since it has the highest average probability. Other responses by subjects will emerge due to different assessments (by different subjects) of segment-pairs at the first stage of decision.

---

[7] This is motivated by the fact that the two interactions found in Chapter 5 (KS and SP) both involve silent portions; perhaps subjects treat silence differently than speech sounds.

The Segment Allophone Model

This model is a special case of the PE model. Whereas the PE model used pairs of segments, the SA model uses the segment as a minimal unit. This segment unit can be considered allophonic in nature and therefore, allophones of the segment are considered as the minimal unit in this model. For example, in the perception experiments in Chapter 5, the K could be considered as a short K or a long K. These decisions are independent of surrounding context. The decisions that such allophones exist would lead to tentative transcriptions. For example, K: would signal K#, or S: would signal #S. These allophones thus cue word beginnings (or endings) as a function of their distribution in natural production (i.e., S: does not signal S# since word-final S's in English are not normally long; see Chapter 3). These transcriptions are stored in memory; when a sufficient number of them are stored, a juncture decision is made, based on the confirmatory information provided by the tentative transcriptions. In this case, where all cues confirm one another, the decision is easy; subjects choose the juncture location that all confirmatory cues signal. An illustration is shown below.

Example 2: The SA Model with No Conflict in Cues

The subject can decide whether he hears a long or short K (K: or K) a long or short S (S: or S) a pause (P) or

no pause (0) and a long or short N (N: or N). If there is no conflict in cues subjects may assess the durations of these elements as follows:

K, a fusion cue which suggests that the K is not at word beginning or ending[K-]

S, a fission cue which suggests that the S signals the end of a word[-s#]

P, is a fission cue [#]

N:, a fission cue which suggests that the N signals the beginning of a word. [#n]

Notice that the pause assessment is a special case, since its distribution in natural speech is, in many cases, 'optional'. This reflects the KS and SP interactions found in the perception experiments, above; silence perhaps constitutes a signalling element that is fundamentally different from other speech sounds.

These allophone decisions give the following tentative transcription:

[K-][-S#][#][#N]

where - represents a fusion cue and #, a fission cue. These markers, - and #, will be called 'segmentation markers'. Compatable segmentation markers can be pruned, yielding

K-S#N

Since all cues combine to produce a segmentation response of LIKES#NAPPING, the subject will so respond.

However, the subject cannot always tell with 100%

accuracy whether the elements are indeed 'long' vs. 'short'. Since these are durational continua involved, it is more appropriate to designate probabilities of assessments. An example is given below. Suppose that the subject assesses these elements with the following probabilities:

K:(.5) S:(.8) 0(.6) N-(.9)

K-(.5) S-(.2) P(.4) N:(.1)

This would suggest that the subject is unsure about the length of the K, fairly unsure of the existence of the pause, and more confident of the S and N durations. Such an assessment will produce several segmentation possibilities:

$$\left\{\begin{matrix}[k\#] \\ [k-]\end{matrix}\right\}\left\{\begin{matrix}[\#s] \\ [-s]\end{matrix}\right\}\left\{\begin{matrix}[P] \\ [0]\end{matrix}\right\}\left\{\begin{matrix}[\#n] \\ [-n]\end{matrix}\right\}$$

where all possible pathways are checked. Other segmentations are not possible since they are never produced in English (eg. LI#KSNAPPING). Furthermore, such segmentations would be involved in the duration assessments of other segments, notably the vowel AI in LIKE.

These possibilities are associated with their elements' respective probabilities as shown above. A tentative transcription is produced depending on the highest average probability combining all relevant cues. In this example, two tentative transcriptions are kept, since they have the same (highest) average probability.

Therefore, the LIKE#SNAPPING and LIKES#SNAPPING responses will be chosen equally, since their average probabilities are equal and they add up to a higher level than any other possibility. In this particular case, higher level processing would be required in order to sort out whether the signal intended one S or two S's and in the categorization experiments conducted in Experiment 4, this would lead to an equal number of S#S and K#S responses. As in the PE model, variable responses are accounted for by suggesting that there are differences among subjects in the first stage allophone decision.

**Conflict in Cues**

When one cue conflicts with another cue in the signal, this conflict must be resolved. This resolution is accomplished by adhering to phonotactic constraints - a type of top-down correcting mechanism. When there is a conflict in cues, speakers have not enunciated as clearly as possible, thereby producing 'noise' in the signal. This makes communication more difficult, since the listener must decode the information from this noisy signal. This is done by bottom-up processing of the signal and a top-down 'check' that will prune out conflicts of the signal. These conflicts are a function of the language structure; phonotactic constraints are broken in the case of conflicting cues.

Thus, listeners should show markedly longer reaction times in these cases than when no cues conflict. In a rating

task, listeners should rate these utterances lower than utterances with no conflict in cues.

The Pairwise-evaluation Model

In the PE model, these phonotactic constraints would outline possible combinations of segment pairs. Recall that the minimal unit is a segment pair which is defined in terms of sound classes (eg., [fricative+stop]). The phonotactic constraints therefore outline the possible combinations of segment pairs. For example, [fricative+stop][vowel+stop]would be a permissable string for English, but [fricative+stop] [fricative+stop] would not. Another example is is where three adjacent elements lead to two segment-pair decisions, i.e., the tentative transcription is [K-S][S-N]. In this case the geminate /s/ condition is signalled (LIKES#SNAPPING), since it is only in this case that both the K-S and S-N are clustered. A third example is the case where there are three adjacent elements and no segment-pair decision is made, since this would suggest that at least one element stands alone. An example of this case is given below.

Example 3: The PE Model with Conflict in Cues

Suppose the pairwise decision was as follows:
K:#S:, which suggests that the K and S are not clustered
S:#N:, which suggests that the S and N are not clustered
This yields a tentative transcription of [K#] [S#][#N]. Since the S cannot stand alone, a decision one way or

the other must be given. In this case, either the LIKES#NAPPING or LIKE#SNAPPING response would be obtained, since in the LIKES#SNAPPING segmentation, the S is clustered with *both* K and N elements. Therefore, the double S condition is ruled out. In these cases, the average probability of pairwise assessments could determine which response would eventually be chosen. For example, suppose the subject were more certain that the K and S were clustered than the S and N. Therefore, he would choose the LIKES#NAPPING response to a greater degree. This would be done in the same manner as described above. The pause is treated in the same way as described above: it strengthens the decision that two adjacent sements should not be paired.

The Segment-allophone Model

In the SA model, the phonotactic constraints would deal with permissable segment combinations. These constraints are of a first-order type that deal with the segments themselves, rather than classes of sounds, whereas the PE model has a second-order constraint, dealing with classes of paired signals. The SA model suggests that, for example, [S:][R] is a nonpermissable string whereas [S:][W] is permissable. Some examples of the SA model in the case of conflicting cues is given below.

Example 4: The SA Model with Conflict in Cues

Suppose the subject assesses the durations as follows:

K, a fusion cue which suggests that K is not at a word beginning or ending.

S:, a fission cue which suggest that the S signals the beginning of a word

#, a fission cue

N:, a fission cue which suggests that the N signals the beginning of a word.

This yields a tentative transcription of:

[-k][#s][#][#n]

Since there is only one instance of like segmentation markers (i.e.,[#][#n]), the segmentation which is cued by this string combination is more likely to be chosen (i.e., the LIKES#NAPPING response). Thus, phonotactic constraints 'correct' the placement of S in the signal by changing the S from [s#] to [#s]. This is done in order to comply with the phonotactic constraint that elements cannot stand alone (i.e., *[#s][#n]). The other half of the transcription ([k-][#s]) would suggest the geminate S case, since the K is short and S is long and would thus comply with the phonotactics only if the S#S case is signalled (i.e., this is a special phonotactic rule). Both phonotactic constraints correct the conflict by attaching the S to the previous word (i.e.,to LIKES rather then to SNAPPING).

The worst case condition, when many cues conflict, is when the tentative transcription reads

[k#][#s][#][#n]

Again, phonotactic constraints must correct these conflicts. This is done by assessing the probability strengths of each element and choosing that boundary segmentation which has the highest average probability. The crossed experiments in Chapter 5 showed that, for this case, later cues dominated for the listeners (see Fig 9 where all elements are long). This suggests that, for this case, the phonotactic constraints pass from right to left as follows:

[#][#N] $\longrightarrow$ [#N]

*[#S][#] changes to [-S#][#]

*[K#][#S] changes to [K-][-S#]

Thus, earlier cues are changed so as comply with later cues.

Of course, the probabilities associated with these decisions are important in all these conflicting cue cases. The output could change depending upon the probability strength of the elements involved.

## C. Discussion

The major differences between models of speech perception, as mentioned above, lies in the choice of relevant units and the nature of the 'decision' or labelling process. Proposed models have suggested that perceptual

units are contextual allophones (Wickelgren, 1969), transitional allophones (Klatt, 1979), allophones in the descriptive linguistic sense (Nearey and Hogan, 1980, and the SA model above), featural (Oden and Massaro, 1979), segment pairs (the PE model above), and finally, the word (Klatt, 1979). What is needed, therefore, are experimental techniques designed to test these different models.

It may prove to be very difficult, if not impossible, to determine the 'basic unit' of speech perception. Licklider (1952) pointed out that certain types of feature-based and template-based systems are functionally equivalent. However, if a unit is postulated in a model, it must be shown to have some psycholinguistic importance. Thus, it should be shown to have an effective influence in speech perception in a linguistically relevant task. In the experiments of Chapters 4 and 5, it is clear that duration is influencing the listeners in a linguistically relevant way - alternative word segmentations of the speech signal emerge when durational patterns are altered. However, these experiments do not point to the basic unit involved. While it is known that duration is important, it is not known what constitutes the decision unit. Is it duration of the segment or is it the fact that duration influences perception of clustering (or segment-pairs)? Perhaps the duration values prime word segmentation based on syllable or word units. Experiments dealing with these types of questions are needed.

For example, if the segment-pair is suggested as a unit, this unit must display some psychological importance when duration is held constant. For example, suppose the utterance "The sheep like(s) srapping (wrapping)" was tested. If the segment-pair is a minimal unit, it would be expected that subjects consistently report hearing LIKES#WRAPPING under all duration levels of K *found in natural production*. Of course, if this K is lengthened to an extreme, the LIKE#SRAPPING response should emerge, since the signal would be too strong to be phonotactically 'corrected'. This presumably happens in the perception of nonsense words, or non-English sequences; the signal is attended to wholeheartedly and the phonotactic constraints are discarded. However, under normal listening conditions (i.e., listening to English sounds) the phonotactic constraints may be assumed to be operative when a conflict between the signal cues is present.

It may prove to be impossible in the case of distinguishing features from segments since, if a segment is altered, its features are also altered. However, experimental ingenuity may overcome at least some of these problems.

The point at which choices are made can be explored in greater detail by using alternate experimental paradigms. These include: reaction time experiments, interruption tasks, and partial identification (eg., does the subject hear 'like' or 'likes', 'snapping' or 'napping', a long S or

a short S, a cluster or a non-cluster?). The choice of the experimental paradigm may prove to be very important. For example, in the the fully crossed experiment in Chapter 5, it was found that the response preference found in the other experiments toward a K#S word segmentation was not apparent. Perhaps this experiment was too structured and allowed for ad hoc strategies to develop (eg., monitoring S durations). A variety of stimuli, including distractor stimuli, is needed so that these types of problems will be reduced.

If a 'knowledge' of natural distribution in speech is acquired, adult subjects should be able to rate utterances as to whether they are 'typical' tokens of a category (see Nearey and Hogan, to appear). Such rating experiments could also test whether conflicting cues in the signal lead to lower ratings than for those utterances in which no cues conflict.

A further area of study is the relation between mathematical models, speech perception, and natural distributions of speech (Nearey and Hogan, to appear). The juncture perception models outlined above lead to different mathematical outputs as a result of probabalistic decisions. These could be further refined and tested. An additional suggestion is to change the response task, away from multi-categories, in order to simplifiy the statistical models which must be applied.

**Conclusions**

The aim of research is to open up new ways of looking

at old problems, but the aim of experimental research is to give substantive support to these new directions. Further experimentation in the area of speech segmentation may offer this support to the models outlined above. Further modelling may yet open other doors in approaching this complex problem.

The area of speech segmentation is new in that only several investigators have approached the subject explicitly. It is hoped that more interest in the area will develop, and provide more information about this complex issue. In particular, more emphasis is required on the nature of the acoustic signal and how it is handled by listeners of the language. It is for this reason that 'bottom-up' models are proposed. Obviously, a complete model will incorporate all possible processes involved in speech, but it is felt that detailed modelling of each process is required first. This research has been an attempt to outline possible bottom-up models, analyze their problems, and give some support that such models, in general, are involved in the speech perception process.

# Bibliography

Abel, S. 1972. Duration discrimination of noise and tone bursts. *Jour. Acoust. Soc. Am. 51:* 1219-1223.

Anderson, S. 1975. On the interaction of phonological rules of various types. *Jour. of Ling. 11:* 39-62.

Andresen, B. 1979. On the perceptability of morphological couplings in English. Paper presented at the International Congress of Phonetic Sciences, Copenhagen.

Bladon, R. and Al-Bamerni, A. 1976. Coarticulation resistance in English /l/. *Jour. of Phon. 4:* 137-150.

Bloomfield, L. 1933. *Language* New York: Holt, Rinehart and Winston, Inc.

Bock, R. and Jones, L. 1968. *The Measurement and Prediction of Judgement and Choice.* San Fransisco: Holden Day.

Cena, R. 1978. *When is a phonological generalization psychologically real?* Indiana University Linguistics Club.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English* New York: Harper and Row.

Chomsky, N., Halle, M. and Lukoff, F. 1956. On accent and juncture in English. In *For Roman Jakobson.* The Hague: Mouton, 65-80.

Coker, C. and Umeda, N. 1975. The importance of spectral detail in initial-final contrasts of voiced stops. *Jour. of Phon. 1:* 63-68.

Christie, W. 1974. Some cues for syllable juncture perception in English. *Jour. Acoust. Soc. Am. 33* 842(A).

Crowder, R. and Morton, J. 1969. Precategorical acoustic

store (PAS). *Percept. and Psychophys. 5:* 365-373.

Cutting, J. 1975. Aspects of phonological fusion. *Jour. of Expt. Psych: Human Percept. and Performance. 104:* 95-112.

Darwin, C. and Brady, S. 1975. Voicing and juncture in Stop-/r/ S clusters. Unpublished manuscript.

DeMori, R. 1977. Syntactic recognition of speech patterns. In *Syntactic Pattern Recognition Applications.* Ed. K. Fu. New York: Springer-Verlag.

Devine A. and Stephens, L. 1976. The function and status of boundaries in phonology. In *Linguistic Studies Offered to Joseph Greenberg.* Ed. Alphonse Juilland. Saratoga, Calif.: Anma Libri.

Eimas, P., Siqueland, E., Jusczyk, P. and Vigorito, J. 1971. Speech perception in infants. *Science 171:* 303-306.

Fujisaki, H. and Kawashima, T. 1969. On the modes and mechanisms of perception of speech sounds. Paper presented at the 78th meeting of the Acoustical Society of America, San Diego.

Haberman, S. 1978. *Analysis of Qualitative Data. Vol. I, Introductory Topics.* New York: Academic Press.

Hoard, J. 1966. Juncture and syllable structure in English. *Phonetica 15:* 96-109.

Halle, M., Hughes, G. and Radley, J. 1957. Acoustic properties of stop consonants. *Jour. Acoust. Soc. Am. 29:* 107-116.

Harms, R. 1968. *Introduction to Phonological Theory.* New Jersey: Prentice-Hall Inc.

Harris, Z. 1951. *Structural Linguistics.* Chicago: University of Chicago Press.

Hill, A. 1958. *Introduction to Linguistic Structures.* New York: Harcourt, Brace and Co.

Hockett, C. 1955. A manual of phonology. *Internat. Jour. of Am. Ling. 21* Baltimore: Waverly Press Inc.

Hockett, C. 1958. *A Course in Modern Linguistics.* New York: The MacMillan Co.

Hooper, J. 1974. Rule morphologization in Natural Generative phonology. In *Papers from the Parassesion on Natural Phonology.* Eds. A. Bruck, R. Fox, and M. LaGaly. Chicago: Chicago Linguistic Society.

Jones, D. 1931. The 'word' as a phonetic identity. *Le Maitre Phonetique. 36:* 60-65

Klatt, D. 1973. Interaction between two factors that influence vowel duration. *Jour. Acoust. Soc. Am. 54:* 1102-1104.

Klatt, D. 1975. Vowel lengthening is syntactically determined in a connected discourse. *J. of Phon. 3:* 129-140.

Klatt, D. 1979. Speech perception: a model of acoustic-phonetic analysis and lexical access. *J. of Phon. 7,* London: Academic Press.

Lass, R. 1970. Boundaries as obstruents: Old English voicing assimilation and universal strength heirarchies. Indiana: Indiana University Linguistics Circle.

Lehiste, I. 1960. An acoustic-phonetic study of internal open juncture. *Phonetica. Supp. 5.*

Lehiste, I. 1964. Acoustical characteristics of selected English consonants. Indiana University Research Centre in Anthropology, Folklore and Linguistics, Publication 34. Bloomington: Indiana University Press.

Lehiste, I. 1970. *Suprasegmentals.* Cambridge: MIT Press.

Lehiste, I. 1972. The units of speech perception. In *Speech and Cortical Functioning.* Ed. J. Gilbert. New York: Academic Press.

Leopold, W. 1948. German ch. *Lang. 24:* 179-180.

Lewis, J., Daniloff, R. and Hammaberg, R. 1975. Apical coarticulation at juncture boundaries. *Jour. of Phon. 1:* 1-8.

Licklider, S. 1952. On the process of speech perception. *Jour. Acoust. Soc. Am. 24:* 590-594.

Lisker, L. 1965. The English stops after /s/ at word boundary; a three-way contrast. Unpublished manuscript.

Luce, R. 1967. Detection and recognition. In *Handbook of Math. Psych. 1:* 103-189. New York: Wiley.

McCasland, G. 1977. English stops after /s/ at medial word-boundary. *Phonetica 34:* 218-228.

McCawley, J. 1968. The phonological component of a Grammar of Japanese. *Monographs on Linguistic Analysis,* The Hague: Mouton.

Massaro, D. and Oden, G. 1980. Evaluation and integration of acoustic features in speech. *Jour. Acoust. Soc. Am. 67:* 996-1013.

Minifie, F., Hixon, T. and Williams, F. 1973. *Normal Aspects of Speech, Hearing and Language.* New Jersey: Prentice-Hall, Inc.

Morton, J. and Long, J. 1976. Effect of word transitional probability on phoneme identification. *JVLVB 15:* 43-52.

Moulton, W. 1947. Juncture in modern standard German. *Lang. 23:* 212-226.

Nakatani, L. 1979. Allophonic and prosodic cues for parsing speech. Paper presented at the International Congress of

Phonetic Sciences, Copenhagen.

Nakatani, L. and Dukes, K. 1977. Locus of segmental cues for word juncture. *Jour. Acoust. Soc. Am. 62:* 714-719.

Nakatani, L. and Dukes, K. 1979. Allophonic, stress and prosodic cues for word perception. Unpublished manuscript.

Nakatani, L. and Schafer, J. 1978. Hearing 'words' without words: prosodic cues for word perception. *Jour. Acoust. Soc. Am. 63:* 234-245.

Nearey, T. 1978. Phonetic feature systems for vowels. Indiana: Indiana University Linguistics Club.

Nearey, T. and Hogan, J. 1979. Normative study of English initial consonants. *Humanties and Social Science Research* Grant No. SSHRC-410-79-0312.

Nearey, T. and Hogan, J. Phonological contrast in experimental phonetics: relating distributions of measurements of natural data to categorization curves. In preparation.

Nearey, T., Hogan, J. and Roszypal, A. 1979. Speech signals, cues and features. In *Perspectives in Experimental Linguistics.* Ed. G. Prideaux. Amsterdam: John Benjamins, B.V.

Nie, N., Hull, C., Jenkins, J. and Steinbrenner, K. and Bent, D. 1975. *SPSS: Statistical Package for the Social Sciences* (2nd edition). New York: McGraw-Hill.

Oden, G. and Massaro, D. 1978. Integration of featural information in speech perception. *Psych. Rev. 85:* 172-191.

Olive, J. and Nakatani, L. 1974. Rule-synthesis of speech by word concatenation: a first step. *Jour. Acoust. Soc. Am. 55:* 660-666.

Pike, K. 1947. Grammatical prerequisites to phonemic

analysis. *Word 3:* 155-172.

Pisoni, D. 1971. Very brief short-term memory in speech perception. Paper presented at the 82nd meeting of the Acoustical Society of America, Denver.

Port, R. 1979a. The influence of tempo on stop closure duration as a cue for voicing and place. *J. Phonetics. 7:* 45-56.

Reddy, D. 1976. Speech recognition by machine: a review. Proceedings of the *IEEE. 64:* 501-531.

Repp, B., Liberman, A., Eccardt, T., and Pesetsky, D. 1978. Perceptual Integration of acoustic cues for stop fricative and affricate manner. *Jour. of Expt. Psych.: Human Perception and Performance, 4:* 621-637.

Rhodes, R. 1974. Non-phonetic environments in Natural Phonology. In *Papers from the Parasession on Natural Phonology.* Eds. A. Bruck, R. Fox, and M. LaGaly. Chicago: Chicago Linguistic Society.

Sag, I. 1974. The Grassmann's Law ordering pseudo-paradox. *Ling. Inquiry 5:* 591-601.

Selkirk, E. 1974. French liason and the X notation. *Ling. Inquiry 5:* 573-590.

Snedecor, G. and Cochran, W. 1967. *Statistical Methods.* Ames, Iowa: The Iowa State University Press.

Stanley, R. 1973. Boundaries in phonology. In *A Festschrift for Morris Halle.* Eds. S. Anderson and P. Kiparsky. New York: Holt, Rinehart and Winston, Inc.

Stevens, K. 1975. The potential role of property detectors in the perception of consonants. In *Auditory Analysis and Perception of Speech.* Eds. G. Fant andM. Tatham. London: Academic Press.

Stevenson, D. and Stephens, R. 1978b. The Alligator reference manual. Unpublished manuscript.

Strevens, P. 1960. Spectra of fricative noise in human speech. *Lang. and Speech. 3:* 32-49.

Sweet, H. 1913. *Collected Papers of Henry Sweet.* Arranged by H.C. Wyld. Oxford.

Trager, G. and Bloch, B. 1941. The syllabic phonemes of English. *Lang. 17:* 223-246.

Trager, G. and Smith, H. 1957. An outline of English structure. *Studies in Linguistics: Occasional Papers 3* Washington: American Council of Learned Societies.

Trubetskoy, N. 1969. Principles of phonology. Translated by C. Baltaxe. L.A.: U. of California Press.

Umeda, N. and Coker, C. 1975. Subphonemic variations in American English. In *Auditory Analysis and Perception of Speech.* Eds. G. Fant and M. Tatham. London: Academic Press.

Vennemann, T. 1974. Words and syllables in Natural Generative phonology. In *Papers from the Parrasession on Natural Phonology.* Eds. A. Bruck, R. Fox, M. LaGaly. Chicago: Chicago Linguistic Society.

Warren, R. 1970. Perceptual restoration of missing speech sounds. *Science 16:* 392-393.

Wickelgren, W. 1969. Context sensitive coding, associative memory, and serial order in (speech) behavior. *Psych. Rev. 76:* 1-15.

Winer, B. 1971. *Statistical Principles in Experimental Design.* (2nd edition) New York: McGraw-Hill.

Wolf, J. 1976. Speech recognition and understanding. In *Understanding in Digital Pattern Recognition.* Ed. K. Fu. New York: Springer-Verlag.

# APPENDIX 1: Stimulus Sentences

The sheep like sleeping
The sheep likes sleeping
The sheep likes leaping

The sheep like sweeping
The sheep likes sweeping
The sheep likes weeping

The sheep like snapping
The sheep likes snapping
The sheep likes napping

The sheep like smashing
The sheep likes smashing
The sheep likes mashing

APPENDIX 2

## Means and Standard Deviations of Measurements

### Group Means

#### Duration

| Juncture | 'Lai' | Closure | Burst | D.S. | Pause | Consonant | Vowel |
|---|---|---|---|---|---|---|---|
| S#S | 208.10154 | 43.70468 | 17.53750 | 196.92656 | 0.0 | 42.96093 | 119.8999 |
| K#S | 211.60311 | 66.65156 | 19.61093 | 154.57343 | 0.0 | 42.36093 | 118.57343 |
| S#N | 215.61405 | 36.51249 | 15.77500 | 117.50156 | 13.67969 | 71.41561 | 135.10156 |
| TOTAL | 211.77290 | 48.95624 | 17.64114 | 156.33385 | 4.55990 | 52.24583 | 124.52499 |

#### Intensity

| Juncture | 'Lai' | Closure | Burst | D.S. | Pause | Consonant | Vowel |
|---|---|---|---|---|---|---|---|
| S#S | 36.51376 | 1.64794 | 17.74829 | 22.08513 | -60.00000 | 36.21068 | 34.83999 |
| K#S | 36.16844 | -1.14612 | 17.58121 | 21.62190 | -60.00000 | 36.39969 | 35.09222 |
| S#N | 36.41537 | 3.94611 | 17.18246 | 19.98602 | -33.90877 | 34.24547 | 35.36499 |
| TOTAL | 36.36586 | 1.48264 | 17.50399 | 21.23101 | -51.30292 | 35.61861 | 35.09907 |

### Group Standard Deviations

#### Duration

| Juncture | 'Lai' | Closure | Burst | D.S. | Pause | Consonant | Vowel |
|---|---|---|---|---|---|---|---|
| S#S | 32.15765 | 13.43958 | 5.82181 | 22.63583 | 0.0 | 11.55755 | 36.49715 |
| K#S | 32.65495 | 32.13292 | 5.08927 | 17.92052 | 0.0 | 10.10190 | 34.68830 |
| S#N | 34.25451 | 11.01525 | 4.69471 | 18.70120 | 16.86482 | 20.10879 | 34.51117 |
| TOTAL | 33.00485 | 24.62149 | 5.42800 | 38.06233 | 11.64550 | 19.89620 | 35.85564 |

#### Intensity

| Juncture | 'Lai' | Closure | Burst | D.S. | Pause | Consonant | Vowel |
|---|---|---|---|---|---|---|---|
| S#S | 2.07635 | 12.56507 | 3.26143 | 2.96138 | 0.0 | 3.54351 | 4.77943 |
| K#S | 2.21998 | 15.40990 | 3.73396 | 3.26730 | 0.0 | 3.54970 | 4.70884 |
| S#N | 1.93667 | 9.64613 | 3.88803 | 3.23373 | 31.95076 | 3.97191 | 4.89110 |
| TOTAL | 2.07507 | 12.43657 | 3.62631 | 3.26772 | 22.10862 | 3.80199 | 4.77339 |

# APPENDIX 3: Duration and S Intensity Values

| N Duration | S Duration | S Intensity (dB.) | Pause Duration | K Closure Duration |
|---|---|---|---|---|
| SN = 38 | SS = 100 | 15 | PA = 0 | TK3 = 23 |
| TNA = 49 | TS3 = 117 | 15 | P = 13 | K13 = 31 |
| NX = 60 | S23 = 127 | 17 | TP3 = 36 | TK1 = 38 |
| TN3 = 70 | S13 = 148 | 17 | K12 = 45 | K12 = 45 |
| LN = 81 | T52 = 167 | 19 | LP = 72 | MK = 54 |
| | S12 = 178 | 19 | | TK2 = 60 |
| | TS1 = 188 | 19 | | LK2 = 67 |
| | SL = 202 | 19 | | LK = 85 |
| | SX = 255 | 19 | | RLK = 100 |
| | | | | ULK = 115 |

# APPENDIX 4: Segments Used in the Crossed Experiment

| Segment | Segment Name | Duration |
|---|---|---|
| N | SN | 40 |
| | NX | 60 |
| | LN | 80 |
| S | SS | 100 |
| | S23 | 127 |
| | TS2 | 167 |
| | SL | 200 |
| P | PA | 0 |
| | TP | 36 |
| | LP | 72 |
| K | TK3 | 23 |
| | MK | 54 |
| | LK | 85 |

B30287